

REVIEW ARTICLE

Regulation of the Globin Genes

ANTONIO CAO AND PAOLO MOI

Istituto di Clinica e Biologia dell'Età Evolutiva, Università di Cagliari, Cagliari, Italy

ABSTRACT

The α - and β -globin gene clusters are subject to several levels of regulation. They are expressed exclusively in the erythroid cells, only during defined periods of development and in a perfectly tuned way, assuring, at any stage of ontogeny, a correct balance in the availability of α - and β -globin chains for hemoglobin assembling. Such a tight control is dependent on regulatory regions of DNA located either in proximity or at great distances from the globin genes in a region characterized by the presence of several DNase I hypersensitive sites and known as the Locus Control Region. All these sequences exert stimulatory, inhibitory or more complex activities by interacting with transcription factors that bridge these regions of DNA to the RNA polymerase machinery. Many of these factors have now been cloned and the corresponding mouse genes inactivated, shading new light on the metabolic pathways they control. It is increasingly recognized that such factors are organized into hierarchies according to the number of genes and circuits they regulate. Some genes such as GATA-1 and 2 are master regulators that act on large numbers of genes at early stage of differentiation whereas others, like EKLF, stand on the lowest step and control only single or limited number of genes at late stages of differentiation. We will review recent data gathered from

expression studies in cell cultures, in transgenic or K.O. murine models as well as from a clinical settings. We will also discuss the development of novel theories on the regulation of the α - and β -globin genes and clusters. (*Pediatr Res* 51: 415–421, 2002)

Abbreviations

HSC, hematopoietic stem cell
LCR, locus control region
HS, (DNase I) hypersensitive site
Fog, friend of GATA-1
NF-E, nuclear factor erythroid
EKLF, erythroid Kruppel-like factor
FKLF, fetal Kruppel-like factor
SSE, stage selector element
SSP, SSE binding protein
CP2, Caat binding protein 2
COUP-TFII, chicken ovalbumin upstream promoter–transcription factor II
XPD, xeroderma pigmentosum D
ATR-X, α -thalassemia-mental retardation X-linked
SWI/SNF, switching and/or sucrose nonfermentor

HISTORICAL PERSPECTIVE

Perhaps because blood samples are readily available, globin research has always been on the leading edge of the biomedical sciences, appealing not only to hematologists but also to many scientists working in the related fields of genetics, structural biology, and molecular biology. The convergence of different scientific disciplines and leading scientists in one field has stimulated research and led to impressive advances in our scientific knowledge over the

years. In fact, many of the most exciting biomedical discoveries in the 20th century are linked to globin research. Hb was the first complex protein to have the three-dimensional structure solved (1) and globin the first gene to be cloned (2). Sickle cell disease was the first genetic disease to be characterized at the molecular level (3). The first observation of the occurrence of polymorphisms along the DNA was made in association with the globin gene cluster (4) and the discovery was immediately applied to the prenatal diagnosis of sickle cell and thalassemia diseases by DNA analysis (5). The field of gene regulation is also heavily indebted to scientists working on globins. β -globin was the first promoter to be extensively characterized in its constitutive elements (6) and the so-called LCR represented the first demonstration of the existence of distant regulatory elements controlling complex clusters of genes (7).

Received June 26, 2000; accepted September 8, 2000.

Correspondence: Prof. Antonio Cao, Director in Chief, Istituto di Clinica e Biologia dell'Età Evolutiva, Istituto di Ricerca sulle Talassemie e Anemia Mediterranea-CNR, Via Jenner, s/n, 09121, Cagliari, Italy; e-mail: acao@mcweb.unica.it

This work was supported by grants from Assessorato Igiene e Sanità Regione Sardegna: L.R. n. 11 30/4/1990, anno 2000 and by Fondazione Italiana L. Giambone per la guarigione della Talassemia.

REGULATION OF THE B-GLOBIN GENE EXPRESSION

The human β -globin gene cluster consists of five genes arranged in chromosome 11 in the same order in which they are expressed during development: 5'- ϵ -, $G\gamma$ -, $A\gamma$ -, δ -, and β -globin gene (8) (Fig. 1). Physiologically, the expression of the genes is generally regulated in such a way that at any point in development the output of the β -globin-like chains equals that of one of the α -globin chains (α :non- α ratio = 1). This is the result of fine transcriptional, posttranscriptional, and posttranslational control by which the transcriptional excess of the two α -genes is leveled up to the output of the single β -globin gene (9). However, the regulatory mechanisms do not take into account whether the genes, which are yet to be activated, are functional or not. Consequently, when the programmed developmental time for the activation of a globin gene is reached, the transcription machinery will be positioned and engaged on its promoter even when the gene is defective and unable to produce any protein. When the defect resides in the β -globin gene, the decreased β -globin synthesis will result in unbalanced chain production and β -thalassemia.

All of the proximal regulatory sequences required for a correct initiation of transcription, and most of the controls that allow correct developmental regulation in transgenic mice, are present within the first 500 bp, 5' to any globin gene. Interaction with more distant elements is required for maximal expression (7). The promoters of all the globin genes share remarkable homology but they also show unique sequences that may be responsible for conveying the developmental stage specificity of each promoter. In a seminal article, Myers and Maniatis (6) conducted a fine genetic and structural analysis of the mouse β -globin promoter and identified three major regulatory elements, the TATA, CAAT, and CACCC boxes, that were required for full globin gene expression. These elements, with minor sequence variations, are present in all globin promoters. The γ - and β -globin promoters, which, because of their involvement in the fetal to adult globin switching, have been the subject of the most intensive investigation, have a prominent element difference because the γ - shows a duplication of the CAAT box and a single CACCC, whereas the β -globin presents duplicated CACCC boxes and a single CAAT (Fig. 2). These differences may have implications for the developmental regulation of the genes but experimental proof is still lacking. Several ubiquitous factors can bind these sequences in exper-

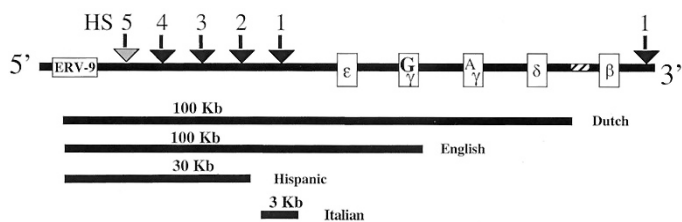


Figure 1. Organization of the β -globin gene cluster. Genes are indicated by open boxes, hypersensitive sites (HS) by vertical arrows, and the origin of replication by a hatched box. The position of the long-terminal repeat of the retrotransposon ERV-9 is indicated upstream of the LCR. The four lower lines represent the extent of the thalassemia deletions of the LCR.

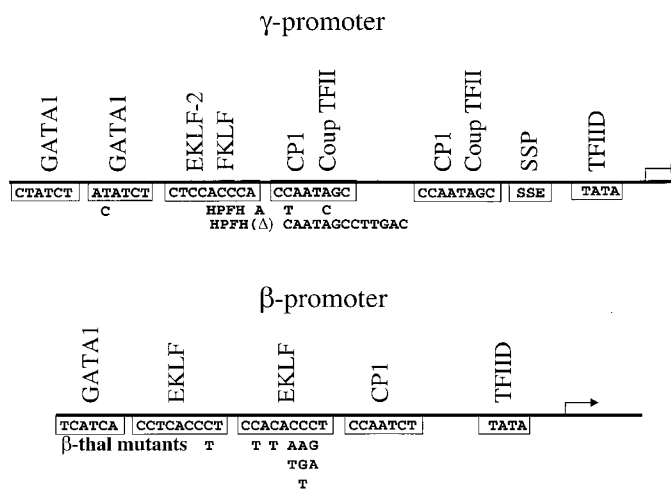


Figure 2. Organization of the γ - and β -globin promoters. The sequences of the most relevant transcription factors binding sites are boxed. Above each site are indicated the factors that are more likely to bind, and below the HPFH or thalassemia point are mutants likely to interfere with transcription factor binding. Note clustering of the HPFH mutants in the γ -globin distal CAAT site and of the thalassemia mutants in the proximal β -globin CACCC box.

imental conditions, but the really determinant factors are more likely erythroid specific factors that have only recently begun to be identified. A sequence element at -50 from the cap site in the γ -globin promoter, SSE, has received much attention as the determinant of the fetal specific activation of the γ -globin gene. In fact, SSE is responsible for the preferential activation of the γ - versus the β -globin promoter when assayed competitively in the same plasmid, and conveys fetal stage expression to the β -globin gene when swapped from the γ - to the β -promoter (10). The embryonic ϵ -globin gene, unique among the β -like globin genes, maintains an appropriate developmental regulation even when associated with the LCR. This behavior is referred to as autonomous developmental regulation to distinguish it from the regulation of the γ - and β -globin genes, which, when linked to the LCR, lose the developmentally regulated expression, unless both genes are simultaneously present in the transgenic construct (competitive developmental regulation). In fact, transgenic mice experiments have established that the developmental control of the ϵ -gene is fully achieved by a DNA fragment of 3.7 kb that includes only the ϵ -gene and 2 kb of upstream sequences, in the absence of any other globin gene in the transgenic construct. Therefore, this fragment should contain all sequence elements required to fully activate and silence the ϵ -globin gene. The activating sequences are ill defined but known to lie within the first 200 bp of the promoter. A silencer has been described further upstream, between -179 and -463 bp, and found to contain DNA binding sites for SP1 or other CACCC binding proteins and for GATA-1. Mutations of the GATA-1 site increase the ϵ -gene expression suggesting that in this context GATA-1 acts as a repressor.

LOCUS CONTROL REGION

DNA sequences in the 5' region of the cluster, up to 50 kb from the β -globin gene, contain four erythroid specific DNase I HS and a further upstream site (HS-5) that is ubiquitously and constitutively on and may signal the boundary of the cluster

(11, 12). HS are a hallmark of DNA-protein interactions and of peculiar modifications or modeling of the chromatin structure (open structure) that facilitate the access of regulators and lower the threshold for activation of the linked genes. Each HS is generally constituted of combinations of several DNA motifs for interacting transcription factors, among which the most important are GATA-1 and NF-E2 (13). Because they confer a position-independent, high level of expression on linked globin genes in transgenic mice (7), the four HS together are referred to as the LCR of the β -globin gene cluster. Since the discovery of the LCR, a great number of HS dissections and combinations have been done in different experimental systems to clarify how the LCR works, but also to identify the minimal sequence requirements to create an LCR compact enough to fit available vectors for the gene therapy of the thalassemias and sickle cell disease. Despite differing opinions and often-contrasting models proposed to explain the LCR mode of action, there is a consensus among globin investigators on several points. Individual or combined HS fragments never reach the level of activity of the intact full LCR. Even though some experiments suggested specific interaction and activation among individual HS and globin genes (14), the current favored model suggests that the LCR acts as a holocomplex of all HS that interacts with only one globin gene at a time during erythroid development. Specific gene expression is the result of a dynamic competition for interaction among the genes and the LCR. As anticipated, in the competitive model of developmental regulation, the γ - and β -globin genes are developmentally regulated only when the LCR and both genes are simultaneously present in the transgene construct (15, 16). Otherwise, inappropriate expression occurs for the β -globin gene during fetal stages and for the γ -globin gene during adult stages of erythropoiesis. Besides competition, other variables that affect the strength and time of interaction between LCR and globin genes, and thus the overall expression, are the gene order (17), the distance from the LCR (18) and the transcription factor milieu, which in turn varies with the stage of development. The importance of the transcription factor environment is clearly demonstrated by the disruption of the EKLF murine gene. In the absence of EKLF, DNase I HS does not form in the β -globin promoter and in the LCR region, and the γ - to β -globin switch is disrupted by the persistence of higher γ -globin levels and severe reduction of the β -globin chains (19–21). The interaction of the LCR with a gene and thus its activation is not irreversible. In fact, primary γ -globin transcripts can be found in cells that have already switched and display predominant β -globin chain in the cytoplasm. This experimental evidence indicates that the LCR can flip-flop back and forth to activate adjacent genes such as the γ - and β -globin genes and that it is the total time spent by the LCR in each gene interaction that determines the prevailing chain in the cytoplasm (22). In experiments where the five β -globin genes or LCR in transgenic mice were inverted in a YAC containing the full β -globin cluster (23), it was determined that the normal orientation of the genes to the LCR is important for developmentally regulated high expression of the globin genes. In fact, inversion of the globin genes so that the β -globin gene is next to the LCR causes inappropriate fetal expression of the

β -globin gene, whereas LCR inversion dramatically decreases the expression of all the globin genes. In the latter configuration, the interposition of HS-5, which may have intrinsic insulating properties (24), between the LCR and the globin genes could explain the generalized suppressive effect (23). Such insulating activity, characterized by enhancer blocking and heterochromatin suppressing effects (25), has clearly been demonstrated in the chicken β -globin cluster for the homologous site (HS-4).

The regulatory role of the LCR is also strongly supported by the silencing of all the β -globin genes produced by the naturally occurring deletions of the LCR in the ($\gamma\delta\beta$)-thalassemias (8) (Fig. 1). Even though deletions of any single HS by homologous recombination in mice had only minimal effects on globin expression, the deletions of the full LCR reproduces in mice the severity of the human ($\gamma\delta\beta$)-thalassemias and is lethal *in utero* (26). However, the normal formation of the HS suggests that other sequences in the human ($\gamma\delta\beta$)-thalassemia deletions are responsible for establishing an open chromatin configuration. A long terminal repeat of the human endogenous retrovirus ERV-9, recently found 5' to HS5 in the human beta globin cluster, has features of erythroid regulatory regions, functions as an erythroid specific enhancer, is evolutionarily conserved and may serve a relevant host function in regulating the transcription of the β -globin (27). ERV-9 generates a large extragenic transcript that originates in its long-terminal repeat TATA box and extends further downstream as far as the ϵ -globin gene. Similar intergenic transcripts are found all over the cluster, exclusively in the erythroid tissues. Although their significance has not yet been established, they are thought to be important in keeping an open configuration of the chromatin (28). Beside ERV-9, several other enhancers have been described in the β -globin gene cluster, the strongest of which is contained in HS2 (29–31).

An origin of replication, described in proximity to the β -globin gene, determines early replication in the erythroid tissues and becomes late replicating when the cluster is silenced as exemplified by the Spanish ($\gamma\delta\beta$)-thalassemia deletion.

Hb SWITCHING

The most intriguing and one of the most studied events that occurs in the β -globin cluster is the so-called Hb switching, the suppression of the γ -globin gene accompanied by the complementary rise in the expression of the previously silent β -globin gene (8). Understanding this process has obvious therapeutic implications for the therapy of β -thalassemia and sickle cell disease. In fact, because the γ -globin gene can functionally substitute for the defective β -globin gene of these diseases (32), any advancement in the comprehension of the Hb switching is likely to benefit therapies based on the reactivation of the γ -globin gene. The regulation of this phenomenon appears to be mainly a matter of promoter regulation as the γ - and β -globin genes are developmentally regulated in transgenic mice, even in the absence of LCR.

Switching factors. The major advances in the field are related to the discovery of novel transcription factors that, by

binding to the promoter elements, regulate the stage-specific expression of the γ - and β -globin genes. One such factor is EKLF (33), which binds the β -CACCC with higher affinity than the slightly different γ -globin CACCC box and preferentially activates the β -globin (34, 35). Furthermore, EKLF is expressed more in the adult stages of development (36) and its gene inactivation in the mice leads to a clear picture of β -thalassemia. Thus, EKLF behaves as an adult switching factor. The search for corresponding fetal specific factors by homology screening has led to the isolation of two more factors, FKLF and FKLF-2. FKLF is more active on the ε - and FKLF-2 on the γ -globin gene. FKLF-2 is also a stronger transactivator as it boosts the expression of the γ -globin gene more than 40 times compared with a six-fold transactivation for FKLF. However, the expression profiles of FKLF (37) and FKLF-2 (38) are not absolutely erythroid specific and their role in switching needs further confirmation in experiments on gene disruption. Another factor that may be implicated in Hb switching is SSP (39, 40), a protein that binds to the SSE of the γ -globin promoter. The element has been shown to confer fetal stage expression on a β -globin promoter and to delay and prolong Hb switching when deleted from transgenic mouse constructs (41). SSE is bound by SSP, a complex heterodimer between the ubiquitous CAAT binding protein CP2 and an erythroid-restricted subunit NF-E4, which has been recently identified (42). The latter appears to confer the SSE binding specificity and the preferential activation of the γ - over the β -globin promoter. Knockout of CP2 (43) produces a silent mice phenotype, probably through compensation by homologous proteins, whereas the knockout of NF-E4, from which more meaningful results are expected, has not yet been done. Two more factors are known to influence the expression of the γ -globin gene and may have a role in Hb switching. One is COUP-TFII (44), a retinoic acid orphan receptor that corresponds to the erythroid specific binding activity of NF-E3 (45, 46). The protein binds to a consensus site that includes the γ -globin CAAT element and may act as a repressor of the γ -globin expression (47). The last putative switching factor is PYR (48, 49), a complex of several proteins with SWI/SNF activity that promote chromatin modification and remodeling (50). PYR is an adult stage-specific factor that binds to a pyrimidine-rich DNA stretch between the γ - and β -globin genes from where it might repress γ -globin and activate β -globin gene expression. Deletion of the PYR element determines a prolonged and delayed switching. The DNA binding component of the PYR complex is the transcription factor Ikaros (51), previously thought to be a lymphoid specific factor.

PHARMACOLOGICAL INDUCTION OF THE γ -GLOBIN GENE

The first demonstration that HbF synthesis could be increased with drugs dates to 1982, when it was observed that 5-azacytidine, a cytotoxic drug, infused intravenously in monkeys stimulated sustained HbF production. The drug was also active in human trials, but its use was soon discontinued due to concerns on the potential carcinogenesis. Hydroxyurea, another cytotoxic drug, was as effective as 5-azacytidine at

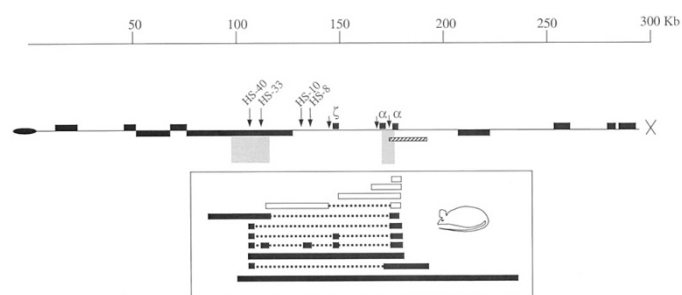


Figure 3. Schematic diagram of the α -globin gene cluster. The oval on the left represents the telomere and the symbol X on the extreme right, the closest site of chromosomal translocation at approximately 1 Mb from the telomere. DNase I hypersensitive sites (HS) are indicated with arrows and the distance in Kb from the ζ -globin gene. Small tail arrows indicate the DNase I sensitive sites in the promoter regions. Globin genes and housekeeping genes are all shown as filled boxes above or below the line according to the direction of transcription toward the centromere or the telomere respectively. Pseudogenes are not shown. Shaded areas indicate the shortest regions of overlap for common α -thalassemia deletions in the α -globin genes and in the regulatory regions. The hatched box below the α -globin genes indicates the extension of the 3' deletion of the cluster that silences the spared $\alpha 2$ -globin gene. The inset summarizes the studies in transgenic mice. The sequences used to assemble the constructs are represented with boxes aligned with the cluster and the sequences not included with dashed lines. The open boxes represent constructs with very low expression (<1% of the mouse genes) and the filled boxes the constructs with good expression. Note that the constructs that express well all bear the HS-40 fragment.

inducing γ -globin synthesis and was preferred in clinical use for the convenience of oral administration and for its better pharmacokinetic and safety profile. Both drugs are thought to act through demethylation of regulatory sequences or by selectively destroying the more mature rapidly cycling erythroid cells and favoring the circulation of earlier progenitors with higher HbF content. The effect of both drugs is transient, variable, and unpredictable. A few years later, it was observed that infants of diabetic mothers had a prolonged switching period that was ascribed to an increase in the blood concentration of butyric acid. Butyric acid and several other short chain fatty acids induce HbF, probably by inhibiting histone deacetylase. Acetylated histones are thought to maintain a chromatin configuration that allows for gene transcription. Also, in selected metabolic diseases, an increase in HbF levels during acute crisis was associated with high plasma levels of short-chain fatty acid metabolites (52). As with hydroxyurea and 5-azacytidine, the clinical response to butyrate is extremely variable and apparently dependent on the genetic background of the patient. Better results have been associated with the intermittent administration of the drug. A promising new field of research correlates the expression of the γ -globin gene with the cellular concentration of cGMP and its repression with increases in cAMP concentration (53). According to this view, general transduction pathways should also be implicated in the regulation of the globin gene expression.

Regulation of the α -globin genes. The α -globin gene cluster resides on the tip of chromosome 16(p13.3) (Fig. 3). The cluster contains three genes and several pseudogenes arranged from the telomere toward the centromere in the following order: 5'- $\zeta 2$ - $\psi \zeta 1$ - $\psi \alpha 2$ - $\psi \alpha 1$ - $\alpha 2$ - $\alpha 1$ - $\theta 1$ -3'. Like the β -globin cluster, the embryonic gene ($\zeta 2$) occupies the place nearest to

the 5' region and the adult globin genes ($\alpha 2\text{-}\alpha 1$) the place nearest to the 3' region. However, the α -cluster lacks a gene expressed uniquely in the fetal period and has two fully functional genes ($\alpha 2\text{-}\alpha 1$) that cover fetal and adult expression, combining with the γ -globin gene during fetal erythropoiesis and with the δ - and β -globin genes in the adult stages of development. Even though the globin genes derive from a common ancestor and share structural similarities, the regulation of the α -globin genes differs in many aspects from that of the β -globin gene (54). Competition among genes does not appear to take place in the regulation of the α -globin gene cluster as all the genes are appropriately regulated when individually linked to the β -LCR or the -40 HS element (55). However, like the ϵ -globin gene, the embryonic $\zeta 2$ -gene is silenced in adult life through DNA sequences that lie near the promoter but are distinct from the corresponding ϵ -globin gene silencer sequences (56). Contrary to the interstitial position of the β -globin gene cluster (11p15.5), the α -cluster occupies a telomeric chromosomal position that places the cluster in a highly unstable and variable region and may have implications for overall gene expression. The whole α -cluster region is highly GC rich, on average 54%, and has a constitutive open configuration of the chromatin and high density of adjacent constitutively expressed nonglobin genes. The repetitive sequences in the α -globin cluster are of the Alu family, whereas the β -globin gene cluster is rich in LINE repetitive elements. The α -globin gene cluster also differs from the β -cluster in its earlier replication time and in the absence of matrix attachment regions. It is still uncertain if the α -cluster contains sequences with LCR function. A clue to the possible presence of LCR elements in the 5' region of the cluster came from human natural deletions of that region. These deletions, like the $\gamma\delta\beta$ -thalassemia deletions of the β -cluster, removed an HS located at -40 kb from the ζ -gene and abolished the expression of all downstream genes. Moreover, the -40 HS core element displayed the same DNA binding sites (GATA-1, NF-E2, and CACCC binding proteins) of the β -globin cluster HS elements and behaved as an enhancer in transient and stable expression assays. Because of these features, HS-40 was thought to correspond to the β -globin LCR. However, when studied in transgenic mice, the -40 HS was expressed at high levels (54) (Fig. 3) but failed to show any LCR activity. In fact, the expression of the transgenes bearing that site was variable and dependent on the site of chromatin integration, was not copy number dependent, declined with time, and never reached a level of expression relative to the endogenous mouse globin genes that was comparable to those attained by the β -globin LCR. Complementing LCR functions were not present in adjacent sequences up to an extension of hundreds of kilobase in PAC transgenes bearing the full telomeric region of the short tip of chromosome 16. Interestingly, a natural deletion of the cluster that occurred in the 3' region but spared the $\alpha 2$ -gene was found to completely inactivate the structurally normal α -globin gene, raising the possibility that the α -LCR could be located in the 3' region of the cluster. However, further work has established that there is no LCR activity in the deleted region and that the likely explanation for the lack of expression is the spread of heterochromatin and DNA methylation from

adjacent Alu sequences into the juxtaposed α -globin gene. Such a mechanism of gene inactivation has never been observed in the deletions of the β -globin gene cluster. Thus, in spite of the observation that in a human chromosomal translocation, in which the α -cluster is displaced in pieces of at least 1 Mb of chromosome, the α -globin gene expression in the derived chromosome is normal, the experimental evidence accumulated so far does not yet support the existence of an α -globin LCR. Recently, fluorescence *in situ* hybridization experiments that made it possible to determine the distance of the locus from the centromere during the cell cycle demonstrated another significant difference between the α - and β -locus. In erythroid cells, the β -globin cluster has been shown to overlap the centromere and to move away from it while changing from an inactive to an active state of expression. In nonerythroid cells, the β -locus permanently resides in close proximity to the centromere. Signals generated from the α -globin cluster, on the other hand, never overlap the centromere and the expression is always active, independent of the cell cycle phase and of the cell line phenotype (D.R. Higgs, personal communication).

FACTORS AND GLOBIN DISEASES

Even though the majority of the defects leading to β -thalassemia are splicing, frameshift, or nonsense mutations of the β -globin gene, and defects producing α -thalassemia are more or less extended α -globin gene deletions, some human mutations occur within the regulatory regions or in unlinked modifying genes. These mutants will be discussed here because they contribute to an understanding of the mechanisms by which the expression of the globin gene is regulated. Besides the already mentioned large deletions that remove the globin LCR and determine $\gamma\delta\beta$ -thalassemias, many other point mutations are known to affect globin gene expression or produce thalassemia by altering regulatory sequences in the promoter or enhancer regions of the β -globin gene cluster. Some of these mutations fall in regions that are binding sequences or recognition sites for general transcription factors such as the TATA box, the Cap and polyadenylation sites. Interestingly, in the β -globin promoter a cluster of point mutations that determine mild β -thalassemia occur in the β -globin proximal CACCC box, where nine different mutations have been described (57) (for a complete description, see the human globin gene database, available at <http://globin.cse.psu.edu>). The distal CACCC, on the other hand, hosts only one mutation (at position -101 from the Cap site), which is associated with silent β -thalassemia. All the mutations in the proximal and distal CACCCs correspond to invariant nucleotides of the consensus site for EKLF and determine reduced binding and transactivation by (58, 59). No mutations have been reported yet in the variant nucleotides. Thus, the mechanism by which the mutations in the β -globin CACCCs determine silent β -thalassemia is through interference with EKLF function. Also noteworthy is the absence of mutations in the β -globin CAAT box, which suggests that this element is either less important than the CACCC boxes or bound by a repressor.

Unlike β -globin, mutations in the γ -globin usually are associated with a gain of function. In fact, mutations causing lack of function may go undetected because of the limited expression of the gene in the fetus or early intrauterine death. The condition of increased γ -globin expression is generally associated with large cluster deletions or promoter point mutations and is known as hereditary persistence of Hb F (HPFH) (8). Like the previously discussed silent β -thalassemias, HPFH point mutations tend to be clustered in specific promoter regions that are binding sites for known and unrecognized factors with the action of which they interfere (Fig. 2). HPFH from mutations of the distal γ -globin CAAT box are probably the consequence of reduced binding of the recently recognized factor NFE3/COUP-TFII, which, in this setting, appears to act as a repressor. On the other hand, an HPFH mutation at position -175 has been associated with increased binding of the transcription factor GATA-1 (60). Interestingly, the same GATA-1 protein has been shown to bind more tightly to a δ -thalassemia mutant of the 3'-untranslated region than to the wild-type sequence and thus, it is proposed to act as a repressor in that context (61). Three other HPFH genes acting in *trans* to the globin cluster have been mapped on the X chromosome (62) and on the long arm of chromosomes 6 (63) and 8 (64). In the Sardinian population, more HPFH genes exist that do not map to these regions and are the subject of a genome-wide search to pinpoint the corresponding loci and clone the genes. In the same population, β -thalassemia patients with identical genetic mutation (β^{39} nonsense) and homogeneous genotypes in the β -globin cluster display different clinical phenotypes of β -thalassemia (transfusion dependent and independent) (65). This suggests the existence of supplementary HPFH modifying genes that could be identified by expression profile analysis.

An (AT)_x(T)_y polymorphic region at -530 from the β -globin cap site has also been shown to bind a repressor protein, BP1 (66), which decreases transcription of linked β and β^S genes (66, 67). However, other studies have raised the possibility that this polymorphic region might be functionally silent or produce a measurable effect only in condition of erythropoietic stress (68).

Lastly, among the *cis* linked defects it is useful to consider the partial or complete deletions of the β -globin promoter. In fact, these deletions support the current competitive model of the function of the LCR, which predicts that in the absence of the β -globin promoter upstream, genes will have more chances and time for interaction with the LCR and thus will have greater expression. This is confirmed by the unusually high levels of HbA₂ ($>6.5\%$) and HbF in β -globin promoter deletions (69) and by the attenuated thalassemia phenotypes of the fusion chain variant, Hb Kenya, in which the complete deletion of the β -promoter leads to an increase in the expression of the γ/β -globin fusion gene. In the other, more common, fusion variant, Hb Lepore (δ/β) (70), compensation is not sufficient because it depends on a transcriptionally weak δ -globin promoter.

Recently, defects in regulatory genes that affect hematopoiesis and globin gene expression have been reported in human patients. The majority occur in the ATR-X gene (71), which is a transcription factor homologous to SNF2 (72, 73), a helicase-

ATPase protein. Mutations in the ATR-X gene (74) determine α -thalassemia, associated with a severe form of mental retardation and genital anomalies. Conversely, defects that alter the transcription functions of a different helicase, the XPD gene protein, determine a phenotype of trichothiodystrophy and a condition of β -thalassemia (75). The reason why defects in apparently ubiquitous factors affect specifically the expression of the α - or the β -globin gene is not yet known, but is probably due to the differential promoter recruitment mediated by promoter specific interacting proteins.

The other regulatory gene found to host human mutations is GATA-1. A missense mutation in the amino Zn finger (val205met) of GATA-1 occurs in a highly conserved residue that participates in the interaction with Fog-1 (76). The resulting clinical phenotype is not lethal, but rather produces a condition of X-linked dyserythropoiesis with thrombocytopenia and the need for *in utero* transfusion. Thrombocytopenia was not part of the GATA-1 null phenotype, which was early lethal, but it was experimentally produced by a knock-down mutation of the GATA-1 promoter (77). A second mutation within GATA-1, opposite to the docking site for Fog, determines thrombocytopenia and β -thalassemia (78).

Despite all these advances, many missing factors need to be cloned and characterized in their reciprocal interactions and in the hierarchies they establish before the complete fascinating puzzle of the regulation of the globin gene will be solved.

REFERENCES

- Muirhead H, Cox JM, Mazzarella L, Perutz MF 1967 Structure and function of haemoglobin. 3. A three-dimensional Fourier synthesis of human deoxyhaemoglobin at 5.5 Angstrom resolution. *J Mol Biol* 28:117-156
- Rabbitts TH 1976 Bacterial cloning of plasmids carrying copies of rabbit globin messenger RNA. *Nature* 260:221-225
- Ingram V 1957 Gene mutations in human hemoglobin: the chemical difference between normal and sickle cell hemoglobin. *Nature* 180:326-328
- Kan YW, Dozy AM 1978 Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci U S A* 75:5631-5635
- Kan YW, Dozy AM 1978 Antenatal diagnosis of sickle-cell anaemia by D.N.A. analysis of amniotic-fluid cells. *Lancet* 2:910-912
- Myers RM, Tilly K, Maniatis T 1986 Fine structure genetic analysis of a beta-globin promoter. *Science* 232:613-618
- Grosveld F, van AG, Greaves DR, Kollias G 1987 Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51:975-985
- Stamatoyannopoulos G, Grosveld F 2001 Hemoglobin switching. In: Stamatoyannopoulos G, Perlmutter R, Majerus W, Varmus H (eds) *The Molecular Basis of Blood Diseases*. Saunders, Philadelphia, pp 135-165
- Lodish HF, Jacobsen M 1972 Regulation of hemoglobin synthesis. Equal rates of translation and termination of α and β globin chains. *J Biol Chem* 247:3622-3629
- Jane SM, Ney PA, Vanin EF, Gumucio DL, Nienhuis AW 1992 Identification of a stage selector element in the human gamma-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the beta-promoter. *EMBO J* 11:2961-2969
- Li Q, Harju S, Peterson KR 1999 Locus control regions: coming of age at a decade plus. *Trends Genet* 15:403-408
- Grosveld F 1999 Activation by locus control regions? *Curr Opin Genet Dev* 9:152-157
- Lowrey CH, Bodine DM, Nienhuis AW 1992 Mechanism of DNase I hypersensitive site formation within the human globin locus control region. *Proc Natl Acad Sci U S A* 89:1143-1147
- Fraser P, Pruzina S, Antoniou M, Grosveld F 1993 Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev* 7:106-113
- Anderson KP, Lloyd JA, Ponce E, Crable SC, Neumann JC, Lingrel JB 1993 Regulated expression of the human beta globin gene in transgenic mice requires an upstream globin or nonglobin promoter. *Mol Biol Cell* 4:1077-1085
- Grosveld F, de Boer E, Dillon N, Fraser P, Gribnau J, Milot E, Trimbom T, Wijgerde M 1998 The dynamics of globin gene expression and gene therapy vectors. *Semin Hematol* 35:105-111
- 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

18. 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
19. Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F 1995 Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* 375:316–318
20. Perkins AC, Sharpe AH, Orkin SH 1995 Lethal beta-thalassaemia in mice lacking the erythroid CACCC- transcription factor EKLF. *Nature* 375:318–322
21. Perkins AC, Gaensler KM, Orkin SH 1996 Silencing of human fetal globin expression is impaired in the absence of the adult beta-globin gene activator protein EKLF. *Proc Natl Acad Sci U S A* 93:12267–12271
22. Wijgerde M, Grosveld F, Fraser P 1995 Transcription complex stability and chromatin dynamics *in vivo*. *Nature* 377:209–213
23. Tanimoto K, Liu Q, Bungert J, Engel JD 1999 Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature* 398:344–348
24. Li Q, Stamatoyannopoulos G 1994 Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood* 84:1399–1401
25. Bell AC, Felsenfeld G 2000 Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene [see comments]. *Nature* 405:482–485
26. Bender MA, Bulger M, Close J, Groudine M 2000 Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol Cell* 5:387–393
27. Long Q, Bengra C, Li C, Kutlar F, Tuan D 1998 A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta-globin locus control region. *Genomics* 54:542–555
28. Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ 1997 Intergenic transcription and transduction of the human beta-globin locus. *Genes Dev* 11:2494–2509
29. Tuan DY, Solomon WB, London IM, Lee DP 1989 An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A* 86:2554–2558
30. Moi P, Kan YW 1990 Synergistic enhancement of globin gene expression by activator protein-1-like proteins. *Proc Natl Acad Sci U S A* 87:9000–9004
31. Ney PA, Sorrentino BP, McDonagh KT, Nienhuis AW 1990 Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev* 4:993–1006
32. Weatherall DJ, Clegg J 1981 *The Thalassemia Syndromes*. Blackwell Scientific, Oxford, UK
33. Miller IJ, Bieker JJ 1993 A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol* 13:2776–2786
34. Asano H, Stamatoyannopoulos G 1998 Activation of beta-globin promoter by erythroid Kruppel-like factor. *Mol Cell Biol* 18:102–109
35. Donze D, Townes TM, Bieker JJ 1995 Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching. *J Biol Chem* 270:1955–1959
36. Bieker JJ 1996 Isolation, genomic structure, and expression of human erythroid Kruppel-like factor (EKLF). *DNA Cell Biol* 15:347–352
37. Asano H, Li XS, Stamatoyannopoulos G 1999 FKLFL, a novel Kruppel-like factor that activates human embryonic and fetal beta-like globin genes. *Mol Cell Biol* 19:3571–3579
38. Asano H, Li XS, Stamatoyannopoulos G 2000 FKLFL-2: a novel Kruppel-like transcriptional factor that activates globin and other erythroid lineage genes. *Blood* 95:3578–3584
39. Jane SM, Amrolia P, Cunningham JM 1995 Developmental regulation of the human beta-globin cluster. *Aust N Z J Med* 25:865–869
40. Jane SM, Nienhuis AW, Cunningham JM 1995 Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein [published erratum appears in *EMBO J* 1995 Feb 15;14(4):854]. *EMBO J* 14:97–105
41. Ristaldi MS, Drabek D, Gribnau J, Poddie D, Yannoutsos N, Cao A, Grosveld F, Imam AM 2001 The role of the -50 region of the human gamma-globin gene in switching. *Embo J* 20:5242–5249
42. Zhou W, Clouston DR, Wang X, Cerruti L, Cunningham JM, Jane S M 2000 Induction of human fetal globin gene expression by a novel erythroid factor, NF-E4. *Mol Cell Biol* 20:7662–7672
43. Ramamurthy L, Barbour V, Tuckfield A, Clouston DR, Cunningham JM, Jane SM 2001 Targeted disruption of the CP2 gene, a member of the NTF family of transcription factors. *J Biol Chem* 276:7836–7842
44. Ritchie HH, Wang LH, Tsai S, O'Malley BW, Tsai MJ 1990 COUP-TF gene: a structure unique for the steroid/thyroid receptor superfamily. *Nucleic Acids Res* 18:6857–6862
45. Ronchi A, Berry M, Raguz S, Imam A, Yannoutsos N, Ottolenghi S, Grosveld F, Dillon N 1996 Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin. *EMBO J* 15:143–149
46. Ottolenghi S, Mantovani R, Nicolis S, Ronchi A, Giglioli B 1989 DNA sequences regulating human globin gene transcription in nondeletional hereditary persistence of fetal hemoglobin. *Hemoglobin* 13:523–541
47. Filipe A, Li Q, Deveaux S, Godin I, Romeo PH, Stamatoyannopoulos G, Mignotte V 1999 Regulation of embryonic/fetal globin genes by nuclear hormone receptors: a novel perspective on hemoglobin switching. *EMBO J* 18:687–697
48. Vitale M, Di Marzo R, Calzolari R, Acuto S, O'Neill D, Bank A, Maggio A 1994 Evidence for a globin promoter-specific silencer element located upstream of the human delta-globin gene. *Biochem Biophys Res Commun* 204:413–418
49. Acuto S, Urzi G, Schimmenti S, Maggio A, O'Neill D, Bank A 1996 An element upstream from the human delta-globin-encoding gene specifically enhances beta-globin reporter gene expression in murine erythroleukemia cells. *Gene* 168:237–241
50. O'Neill D, Yang J, Erdjument-Bromage H, Bornschlegel K, Tempst P, Bank A 1999 Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. *Proc Natl Acad Sci U S A* 96:349–354
51. Georgopoulos K 1997 Transcription factors required for lymphoid lineage commitment. *Curr Opin Immunol* 9:222–227
52. Little JA, Dempsey NJ, Tuchman M, Ginder GD 1995 Metabolic persistence of fetal hemoglobin. *Blood* 85:1712–1718
53. Ikuta T, Cappellini M 1999 A novel mechanism for fetal globin gene expression: role of the soluble guanylate cyclase-cyclic GMP pathway. *Blood* 94:615
54. Higgs DR, Sharpe JA, Wood WG 1998 Understanding alpha globin gene expression: a step towards effective gene therapy. *Semin Hematol* 35:93–104
55. Albitar M, Katsumata M, Liebhaber SA 1991 Human alpha-globin genes demonstrate autonomous developmental regulation in transgenic mice. *Mol Cell Biol* 11:3786–3794
56. Watt P, Lamb P, Proudfoot NJ 1993 Distinct negative regulation of the human embryonic globin genes zeta and epsilon. *Gene Expr* 3:61–75
57. Huisman T, Carver M, Baysal E 1997 A Syllabus of Thalassemia Mutations. The Sickle Cell Anemia Foundation, Augusta, GA, USA
58. Faustino P, Lavinha J, Marini MG, Moi P 1996 beta-Thalassemia mutation at -90C->T impairs the interaction of the proximal CACCC box with both erythroid and nonerythroid factors [letter]. *Blood* 88:3248–3249
59. Feng WC, Southwood CM, Bieker JJ 1994 Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* 269:1493–1500
60. Martin DI, Tsai SF, Orkin SH 1989 Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature* 338:435–438
61. Moi P, Loudianos G, Lavinha J, Murrus S, Cossu P, Casu R, Oggiano L, Longinotti M, Cao A, Pirastu M 1992 Delta-thalassemia due to a mutation in an erythroid-specific binding protein sequence 3' to the delta-globin gene. *Blood* 79:512–516
62. Miyoshi K, Kaneto Y, Kawai H, Ohchi H, Niki S, Hasegawa K, Shirakami A, Yamano T 1988 X-linked dominant control of F-cells in normal adult life: characterization of the Swiss type as hereditary persistence of fetal hemoglobin regulated dominantly by gene(s) on X chromosome. *Blood* 72:1854–1860
63. Garner C, Mitchell J, Hatzis T, Reittie J, Farrall M, Thein SL 1998 Haplotype mapping of a major quantitative-trait locus for fetal hemoglobin production, on chromosome 6q23. *Am J Hum Genet* 62:1468–1474
64. Thein SL 2000 Identification of factors regulating HbF production: a genetic approach. *Blood Cell Molec Dis* 5:505
65. Cao A, Galanello R, Rosatelli MC 1994 Genotype-phenotype correlations in beta-thalassemias. *Blood Rev* 8:1–12
66. Berg PE, Mittelman M, Elion J, Labie D, Schechter AN 1991 Increased protein binding to a -530 mutation of the human beta-globin gene associated with decreased beta-globin synthesis. *Am J Hematol* 36:42–47
67. Elion J, Berg PE, Lapoumeroulie C, Trabuchet G, Mittelman M, Krishnamoorthy R, Schechter AN, Labie D 1992 DNA sequence variation in a negative control region 5' to the beta-globin gene correlates with the phenotypic expression of the beta s mutation. *Blood* 79:787–92
68. Wong SC, Stoming TA, Efremov GD, Huisman TH 1989 High frequencies of a rearrangement (+ATA; -T) at -530 to the beta-globin gene in different populations indicate the absence of a correlation with a silent beta-thalassemia determinant. *Hemoglobin* 13:1–5
69. Craig JE, Kelly SJ, Barnetson R, Thein SL 1992 Molecular characterization of a novel 10.3 kb deletion causing beta-thalassaemia with unusually high Hb A2. *Br J Haematol* 82:735–744
70. Flavell RA, Kooter JM, De Boer E, Little PF, Williamson R 1978 Analysis of the beta-delta-globin gene loci in normal and Hb Lepore DNA: direct determination of gene linkage and intergene distance. *Cell* 15:25–41
71. Gibbons RJ, Suthers GK, Wilkie AO, Buckle VJ, Higgs DR 1992 X-linked alpha-thalassemia/mental retardation (ATR-X) syndrome: localization to Xq12-q21.31 by X inactivation and linkage analysis. *Am J Hum Genet* 51:1136–1149
72. Picketts DJ, Higgs DR, Bachoo S, Blake DJ, Quarrell OW, Gibbons RJ 1996 ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome. *Hum Mol Genet* 5:1899–1907
73. Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H, Higgs DR 2000 Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* 24:368–371
74. Gibbons RJ, Picketts DJ, Villard L, Higgs DR 1995 Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* 80:837–845
75. Viprakasit V, Gibbons RJ, Broughton BC, Tolmie JL, Brown D, Lunt P, Winter RM, Marinoni S, Stefanini M, Brueton L, Lehmann AR, Higgs DR 2001 Mutations in the general transcription factor TFIIB result in beta-thalassaemia in individuals with trichothiodystrophy. *Hum Mol Genet* 10:2797–802
76. Nichols KE, Crispino JD, Poncz M, White JG, Orkin SH, Maris JM, Weiss M J 2000 Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet* 24:266–270
77. Shivdasani RA, Fujiwara Y, McDevitt MA, Orkin SH 1997 A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J* 16:3965–3973
78. Orkin S 2000 Modulation of GATA-factor activity by Fog cofactors. *Blood Cell Molec Dis* 5:501