# Chromatin dynamics
# and
# the regulation of β-globin gene expression

Mark Wijgerde

# Chromatin dynamics and the regulation of β-globin gene expression

De dynamiek van chromatine en de regulatie van β-globine gen expressie

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof.dr P.W.C. Akkermans M.A.
en volgens besluit van het College van Promoties

De openbare verdediging zal plaatsvinden op
woensdag 10 juni 1998 om 9.45 uur

Door

Marcus Govert Johannes Maria Wijgerde

geboren te Breda

**PROMOTIECOMMISSIE**

| | |
|---|---|
| Promotor: | Prof.dr F.G. Grosveld |
| Overige leden: | Prof.dr D. Bootsma.<br>Prof.dr A. Grootegoed<br>Prof.dr B. Wieringa |
| Co-promotor : | Dr P. Fraser |

Dit proefschrift werd bewerkt binnen de vakgroep Celbiologie en Genetica, Faculteit der Geneeskunde en Gezondheidswetenschappen, Erasmus Universiteit Rotterdam. De vakgroep maakt deel uit van het Medisch Genetisch Centrum Zuid-West Nederland. Het onderzoek werd financieel gesteund door NWO.

Print: Offsetdrukkerij Ridderprint B.V., Ridderkerk

'We must not surmise or invent,
but discover, what Nature does,
as Bacon very well says.'
Swammerdam

Voor mijn ouders.

# Contents

**List of abbreviation**

| | |
|---|---|
| A | Adenine, base |
| bp | Base pair |
| C | Cytosine, base |
| cDNA | Complementary DNA, DNA made by reverse transcribing poly-A positive cellular RNA |
| DNA | Deoxyribonucleic acid |
| DRB | 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole |
| EKLF | Erythroid Krüppel-like factor |
| G | Guanine, base |
| gDNA | Genomic DNA, DNA belonging to the genetic information of a cell or organism |
| GTF | General transcription factor |
| H1 | Histon type 1 |
| H2A | Histon type 2A |
| H2B | Histon type 2B |
| H3 | Histon type 3 |
| H4 | Histon type 4 |
| HAT | Histone acetyl transferase |
| HbA | Adult haemoglobin |
| HbF | Foetal haemoglobin |
| HEL | Human erythroid leukemia (cell line) |
| HPFH | Heriditary persistence of fetal haemoglobin |
| HS-40 | Hypersensitive site -40 |
| HSS | Hypersensitive sites |
| IVS | Intervening sequence, intron |
| kb | Kilo base |
| kD | Kilo Dalton |
| LCR | Locus control region |
| MAR | Matrix associated region |
| MB | Mega Base ($10^6$ nucleotides) |
| MEL | Mouse erythroleukemia line, an immortal cell line frozen in the proerythroblast stage of its differentiation. |
| mRNA | messenger ribonucleic acid |
| NF1 | Nuclear factor 1 |
| PA | Polyadenylic acid |
| PIC | Pre-initiation complex |
| RNA | Ribonucleic acid |
| SAR | Scaffold attachment region |
| T | Thymine, base |
| TBP | TATA binding protein |
| TFIID | Transcription factor II D |
| TK | Thymidine kinase |
| U | Uracil, base |
| UTR | Untranslated region |

| Greek letters | | Symbol | | Factor |
|---|---|---|---|---|
| α | alfa | T | tera | $10^{12}$ |
| β | beta | G | giga | $10^{9}$ |
| γ | gamma | M | mega | $10^{6}$ |
| δ | delta | k | kilo | $10^{3}$ |
| ε | epsilon | m | milli | $10^{-3}$ |
| ζ | zeta | μ | micro | $10^{-6}$ |
| θ | theta | n | nano | $10^{-9}$ |
| ψ | psi | p | pico | $10^{-12}$ |

## Scope of the thesis

The human β-globin locus is frequently used as a model system to study mechanisms controlling tissue-specific and developmentally regulated gene expression. Much of the recent progress in understanding the regulation of β-globin gene expression has come from a better knowledge of the process of transcription. Proper transcriptional regulation of the human β-globin genes occurs, at least in part, through specific interactions of regulatory *trans*-acting proteins to defined *cis*-regulatory sequences that include promoters, enhancers, silencers, and elements of the locus control region (LCR, Chapter 1).

Several models have been proposed for the mechanism of LCR activation and developmental regulation of the β-globin genes. One model is the accessibility model, which suggests that the LCR merely generates topological changes in the chromatin structure making all the genes accessible to freely diffusable transacting factors including the transcription machinery. A second model suggests that the LCR functions as a nucleation site for some transcriptional activation activity or complex, which scans along the DNA to activate gene transcription. The third model is the looping model, which suggests that the LCR participates in direct chromatin interactions with promoter regions of a gene to activate transcription. In trying to discriminate which mechanism is most likely to operate *in vivo* in the regulation of the human β-globin genes, we focused our attention on stages of globin gene switching in order to gain insights into the transcriptional activation dynamics and to make inferences about the mechanism operating in the locus. In Chapter 2 we have developed and applied a sensitive fluorescent *in situ* hybridisation technique to detect primary RNA transcripts at the site of transcription, which allowed us to discriminate between a single and a multiple gene activation mechanism operating in the locus. The results showed that single gene signals dominate and suggested that gene switching is a dynamic process. Since the results were not fully decisive, we used RNA FISH in combination with DRB-induced (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole) transcriptional inhibition and release experiments (Chapter 5) to extend our insights gained into the kinetics of globin gene transcription. The results imply a dynamic, single gene activation mechanism in which gene transcription is an all or none event. We find the results most compatible with the looping model of LCR driven gene activation. In addition, the results predict that globin gene switching is brought about by gradual changes in the *trans*-acting factor environment. This hypothesis was tested in experiments described in Chapter 3, where we show that changes in globin gene expression during periods of γ- to β-gene switching are a result of changes in the concentration of the erythroid Krüppel-like factor (EKLF). Expression studies were performed using compound EKLF knockout and single copy human β-globin locus transgenic mice. This analysis was extended as described in Chapter 4, by studying single copy human β-globin locus transgenic mice in an EKLF overexpressing background. Our data indicates that the concentrations of a critical transcription factor can play an important role in the balance of gene transcription during stages of globin gene switching. In addition we show that EKLF is functional in the primitive lineage by showing that it binds to the CACC motifs in LCR HS site3. We demonstrate that expression of the human adult β-gene is not dependent on EKLF in the embryonic lineage but acquires this dependency in the definitive lineage. How this dependency is acquired is not known.

Each chapter contains a comprehensive discussion that deals with the newly obtained results trying to match them with one of the proposed models. In Chapter 6, I have shortly summarised the results, predominantly in the light of the prevailing looping model. Further, I have discussed some aspect of the RNA FISH technique and some conflicting data concerning the function of EKLF in the primitive and definitive lineages with regard to the activation of the adult β-globin genes.

# Chapter 1

General introduction

# General Introduction

## 1.1 The erythroid lineage, haemoglobin and the globin genes

*Haematopoiesis and the red cell lineage*

Haematopoiesis is the generation of the different blood cell types from a limited pool of multipotential self-renewing stem cells, which as a continuous process is essential throughout the entire life span of vertebrates (Metcalf, 1989). Differentiation into the various lineages is in part specified by interactions of extracellular signals (called haemapoietic growth factors) to membrane bound receptors. After ligand binding and through still ill defined secondary intracellular signalling pathways, these receptors can invoke differential responses in gene expression in pluripotent precursor cells. This results in irreversibly programming these precursors to a defined blood cell lineage such as red cells, neutrophils, monocytes/macrophages, megakaryocytes, mast cells and lymphoid cells (Metcalf, 1989; Olsson et al, 1992; Figure 1).

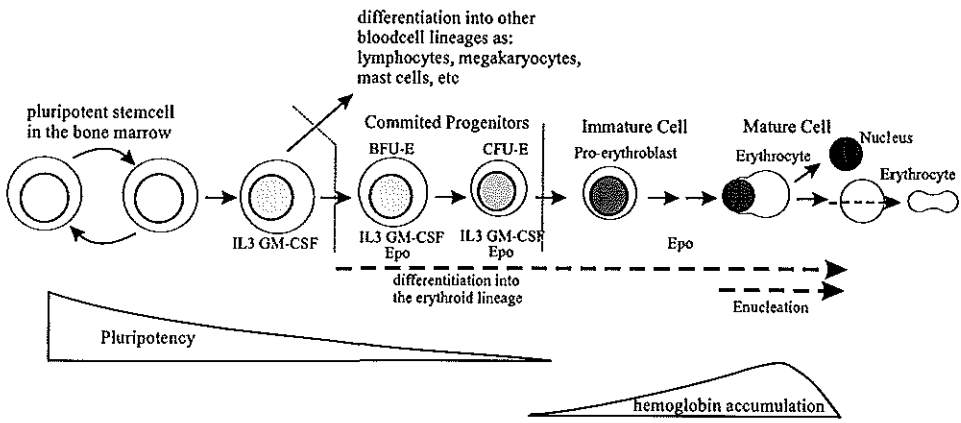



**Figure 1.** Pathway of erythroid maturation.
Some of the growth factors associated with each step in the differentiation pathway are shown. The proliferative potential of the cells decreases with each differentiation step, which is indicated by the bar on the bottom left. The nucleus tends to decrease in volume due to a higher compaction of the chromatin, which is indicated by the smaller size and darker staining of the circles representing the nucleus. Globin mRNA and protein accumulate during the terminal phase of erythroid differentiation, as is indicated by the bar on the bottom right. Abbreviations: BFU-E, burst forming unit erythroid; CFU-E, colony forming unit erythroid. Growth factors: Epo, erythropoietin; IL3, interleukin 3; GM-CSF, granulocyte macrophage colony stimulating factor.

The most common type of cell in the adult mammalian blood is the erythrocyte or red blood cell. Erythroid differentiation takes place in a series of intermediate precursors that progressively gain erythroid features and lose proliferative capacity (Figure 1). At the terminal stages of differentiation, the red cell is deprived of its nucleus, endoplasmic reticulum and mitochondria, making it unable to grow and proliferate. In addition, human erythrocytes have a limited life span of about 120 days. This demands a continuous supply of red blood cells (estimated to be roughly 1-2 million cells per second) through the differentiation of single haematopoietic stem cells. The erythroid compartment represents a well-characterised and accessible system and is therefore frequently used as a model for studying and understanding the mechanisms involved in differentiation and development. During erythroid differentiation a programmed set of hierarchically expressed genes characterise each step in the differentiation process. Among the last genes to become active are those that encode the globin proteins which are exclusively expressed and accumulate to very high levels in terminally differentiated red cells (see Figure 1). In mammals two families of globin polypeptides are distinguished, namely α- and β-globin, which form part of the larger oxygen transporter molecule haemoglobin. During human development red blood cells express different α- and β-like globin peptides which are encoded by a family of α- and β-like globin genes clustered within two different chromosomal loci. These loci have long provided unique model systems for the study of the regulation of gene expression in vertebrates.

*Haemoglobin and haemoglobinopathies*

The contents of circulating red blood cells consist of approximately 90% of the oxygen transporter molecule haemoglobin, providing the red blood cell with its major function, the transport of oxygen from the lungs to all peripheral body cells and the transport of carbon-dioxide from peripheral tissues to the lungs (reviewed in Dickerson, 1983). Haemoglobin is a compound tetrameric molecule composed of two α- and two β-like globin polypeptide chains (Perutz, 1960). Each globin polypeptide chain is associated with a heme molecule, which is incorporated inside a pocket like-cavity formed by three alpha helices. The heme group is an inorganic molecule composed of a divalent iron incorporated within a protoporphyrin IX ring, which is capable of reversibly binding one molecule of oxygen and forms the functional centre of the molecule. Although the heme group is the same in the different globin subtypes produced during human ontogeny, there are distinct variations in the oxygen binding characteristics of the different haemoglobin subtypes that reflect the oxygen requirements of the developing embryo.

Co-ordinated expression of three α- and five β-like globin genes produce the different embryonic, foetal and adult haemoglobin subtypes, which results in a balanced expression of α-like versus β-like globin proteins throughout development. Many genetic abnormalities can influence the haemoglobin production during development or the adult stage, giving rise to the haemoglobinopathies. We can distinguish disturbances in the timing of gene switching during development and disturbances of the level of globin gene expression. Hereditary persistence of foetal haemoglobin (HPFH) is a disorder which affects the developmental switch from foetal γ- to adult β-globin gene expression. HPFH leads to elevated levels of foetal haemoglobin (HbF) in adult life but has no harmful effects, because HbF can adequately substitute for adult haemoglobin (HbA) in β-globin deficient conditions such as β-thalassaemias and is even beneficial in patients heterozygous for the sickle cell allele (Kaul et al., 1996; Wood, 1993; Craig et al., 1996). Sickle cell disease is caused by a single base pair (bp) mutation in codon 6 of the human β-globin gene resulting in a single amino acid substitution of glutamine to valine. The mutant β-globin molecule is susceptible to polymerisation when deoxygenated, causing a number of red cell deficiencies. The red cells deform into a sickle shaped cell vulnerable to degradation and severe anaemia. Sickled cells

also lead to vaso occlusion and multi-organ damage by necrosis. Thalassaemias are caused by perturbation in the production of either α- or β-type globin proteins, resulting in unbalanced quantities of both gene products. This leads to a decreased level of haemoglobin synthesis and formation of protein precipitates of the excess globin chains. The precipitates make red cells vulnerable to degradation, resulting in severe anaemia (Orkin, 1986; Thein, 1993). Two general classes of thalassaemia are distinguished, dependent on whether an α- or β-type gene is affected causing either α-thalassaemia or β-thalassaemia, respectively. The severity of the anaemia phenotype can range from mild to a lethal anaemia depending on the extent of the disturbance between α- and β-protein levels. The thalassaemias represent one of the most prevalent genetic defects in the world, which are thought to be maintained because of a carrier advantage for malaria infection. These haemoglobinopathies have played a very important role in the study of transcriptional regulation during development and will be quoted when necessary.

*Globin gene evolution and gene structure*

Both globin gene subtypes find their ancestral origin in a duplication event that took place about 450 million years ago, leading to the separation into an α-like and a β-like globin gene (Hardison, 1991). After the initial duplication, the α-like and β-like globin genes were free to evolve independent of each other. In amphibians, the α- and β-like globin genes are tightly linked, but these genes became separated on different chromosomes in the lineage to birds and mammals. In their subsequent evolution, the α-like and β-like globin genes duplicated several more times with individual genes being subjected to sequence divergence, deletions, conversions and retro-transpositions leading to the present-day families of globin gene clusters. Despite the long evolutionary separation of the human α- and β-like globin genes, their polypeptide chains still match in about ~50% of the amino acid positions. For excellent and comprehensive reviews on the evolution or history of haemoglobin and the globin genes, see Efstratiadis et al., 1980; Hardison and Miller, 1993 and Hardison, 1991 and 1996.

Early investigations into globin gene structure led to the discovery that coding sequences of genes were divided into segments called exons, which were interrupted by non-coding intervening sequences (IVS) or introns (Jeffreys and Flavell, 1977; Tilghman et al., 1978). The structure of the globin genes has been well-conserved among many species, and has a compacted structure between 1 and 2 kb in size. Virtually all globin genes have two introns, which are found in equivalent positions in all vertebrate α-like and β-like globin genes. Because the exons were thought to encode for discrete domains of protein structure, it has been suggested that the intron-exon structure of the globin genes may have its origin in the shuffling of exons to produce novel proteins (Gilbert, 1978; Blake, 1978; Souza et al., 1996). In the case of the mammalian β-globin genes, the first intron, which varies between 110-130 bp, is much smaller than the second intron, which varies between 600-900 bp. The intron sequences of the Gγ- and Aγ-globin genes show great similarity and so do the introns of the δ- and β-globin genes, while intron comparison of the ε- to Gγ- and Aγ- to δ- and β-globin genes reveal little sequence similarity. Both exons and introns are transcribed into a single continuous precursor RNA, and the non-coding intron sequences are subsequently removed in the nucleus through RNA splicing. The resulting mature mRNA contains a 5' untranslated region (5'UTR) which is around 50bp, followed by the approximately 600 bp coding sequence, a 3'UTR of about 150 bp, and a polyA tail varying in length between 100 to 200 adenylic acid residues (Figure 2). Both UTRs are thought to play an essential role in increasing the stability and translation efficiency of the mature globin mRNAs (Weis and Liebhaber, 1994; Russel and Liebhaber, 1996). Globin mRNAs encode proteins with sizes between 141 and 146 amino acids.

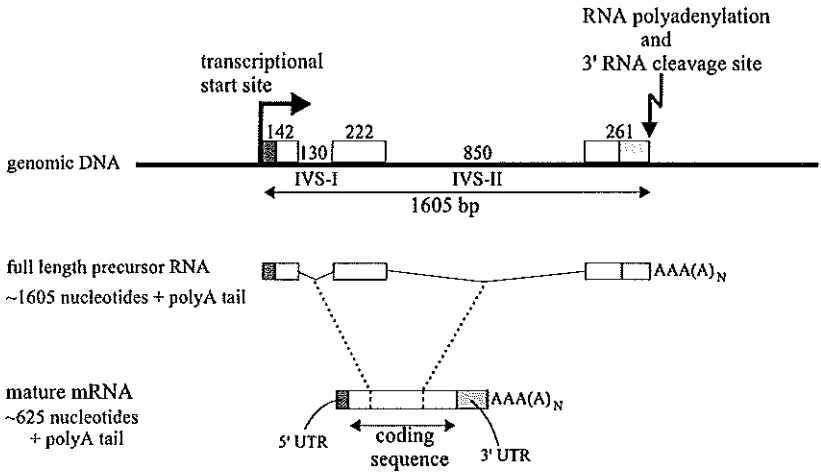



**Figure 2.** Structure of the human β-globin gene, full-length primary RNA transcript and the fully processed mature mRNA.
The white bars represent protein coding regions while the grey bars represent the untranslated 5' and 3' regions (UTR). The numbers indicate the length of intron and exon regions in base pairs (bp). The $AAA(A)_N$ indicates the polyA-tail (pA) which is between 100 and 200 nucleotides (nt) in length.

*Genomic organisation and human globin gene expression during development*

The human α- and β-globin loci are multi gene clusters, which are located on the short arm of chromosome 16 (band 16p13.3) and on the short arm of chromosome 11 (band 11p15.5), respectively. The α-globin locus includes four expressed genes namely the embryonic expressed ζ2-gene and the two foetal and adult expressed α1- and α2-genes and a gene of undetermined function, the θ-gene. No function has been found for the θ-gene and deletions including the entire θ-gene appear to have no recognisable phenotype, still the gene is expressed at very low levels in erythroid tissues (reviewed in Higgs, 1993). In addition, three pseudogenes, ψζ1, ψα2, and ψα1 are embedded in the α-locus. The structure of the α-locus is as follows: 5'ζ2-ψζ1-ψα2-ψα1-α1-α2-θ 3' (reviewed in Higgs et al., 1989; Higgs, 1993; Figure 3). The β-globin cluster spans 70kb and contains a set of five erythroid specific genes including an embryonic gene, ε, two foetal genes, $^{G}\gamma$ and $^{A}\gamma$, two adult genes, δ and β and a non-expressed pseudogene, ψβ. The structure of the β-locus is as follows: 5'ε-Gγ-Aγ-δ-β 3' (reviewed in Collins and Weissman, 1984). In both the human α and β loci the genes are arranged along the chromosome in the same order as they are expressed during development (Figure 3). The α- and β-globin clusters are both preceded by remote *cis*-acting regulatory sequences referred to as hypersensitive site –40 (HS-40) and locus control region (LCR) respectively (Figure 3). HS-40 represents a small region of DNA hypersensitive to DnaseI

treatment which is located 40kb upstream of the embryonic ζ-gene and its presence is essential for the expression of all genes in the α-globin cluster (reviewed in Higgs, 1990). The functional activity of the β-globin LCR is contained within 5 DNA regions characterised by being hypersensitive to DNaseI treatment and which are spread over a 16kb region (Grosveld et al., 1993). Presence of the LCR is important for the expression of all β-genes in the locus and will be more elaborately discussed later.

Early embryonic erythropoiesis takes place in the lining of the yolk sac, where small "islands" filled with primitive nucleated erythroid cells are produced. Around the 3rd week of gestation the α- and β- globin loci are activated in erythroid cells of the primitive lineage, which predominantly express the ε-globin and ζ-globin genes and low levels α- and γ-globin genes. The embryonic haemoglobins produced are of the type Gower I ($\zeta_2\varepsilon_2$), Gower II ($\alpha_2\varepsilon_2$), Portland I ($\zeta_2\gamma_2$) and Portland II ($\zeta_2\gamma_2$). Around the sixth week of gestation erythropoiesis moves to the foetal liver producing anucleated erythrocytes of the definitive lineage, which coincides with a switch from ε- to γ-globin gene expression and from ζ- to α1- and α2-globin gene expression. Around 10 weeks of gestation embryonic haemoglobin is no longer detectable and fetal liver erythropoietic cells mainly produce fetal haemoglobin (HbF, $\alpha_2\gamma_2$). The two α-globin genes remain expressed throughout the rest of development and in adult stages. During the fetal period, the site of erythropoiesis gradually switches from the liver to the spleen and finally the bone marrow. γ-Globin gene expression decreases during this period with reciprocal increases in the expression of the adult β-globin gene and the appearance of low levels of adult δ-globin gene expression. Throughout adult life the bone marrow remains the major site of erythropoiesis with β-globin (~97%) being the major gene and δ-globin (~3%) being the minor gene expressed, producing major (HbA, $\alpha_2\beta_2$) and minor (HbA2, $\alpha_2\delta_2$) adult haemoglobin subtypes respectively (Weatherall and Clegg, 1981; Peschle et al., 1985; Bunn and Forget, 1986; Stamatoyannopoulos et al., 1987 and 1994).

The human α-globin locus is characterised by a single switch in gene expression, while the human β-globin locus is characterised by two major gene expression switches. These major gene-switching events seem to coincide with developmental switches in the site of erythropoiesis (Figure 3). It should be noted however that globin gene expression during the different developmental stages in human is not strictly related to the site of erythropoiesis since ζ-, ε- and β-gene expression can be detected during foetal stages and γ-gene expression can be detected in embryonic and adult erythrocytes.

Erythropoiesis provides a relatively well-defined and accessible model system for the study of differentiation and gene expression during development. How the globin genes restrict their activity to the erythroid lineage and how developmental gene switching is regulated are among the basic problems that we study. Important parameters that have been implicated in the regulation of the globin genes are chromatin structure, transcription factors and *cis*-regulatory DNA sequences, all of which will be discussed.

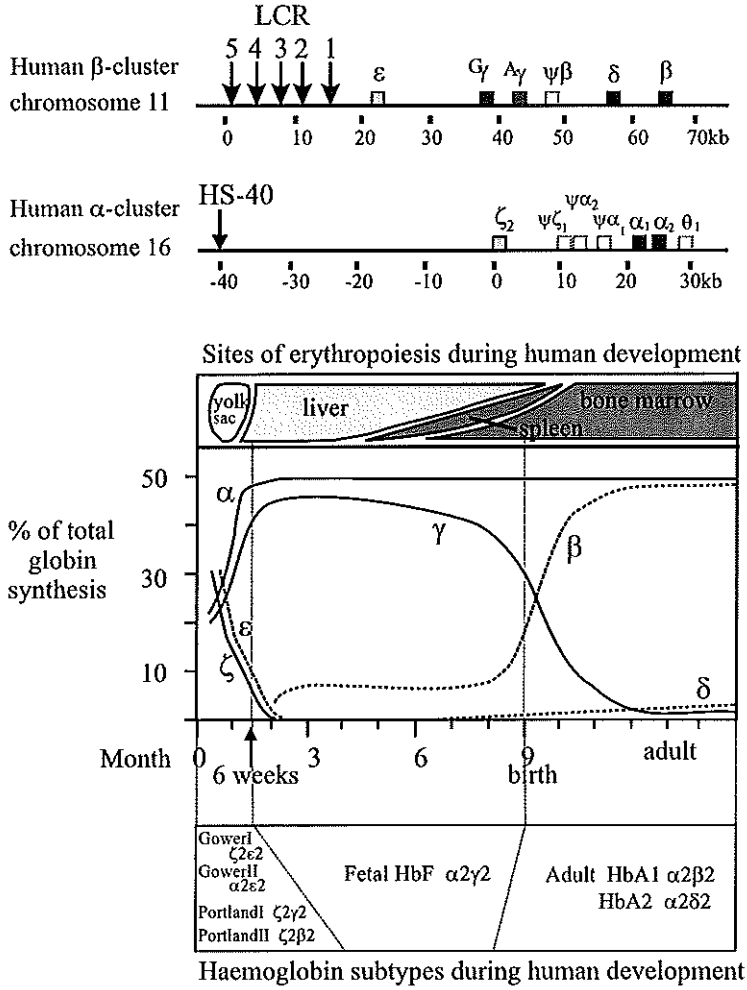



**Figure 3.** Organisation and developmental patterns of expression of the human α- and β-globin genes.
Top: The organisation of the human α- and β-globin locus on chromosome 16 and 11 respectively. Middle: The pattern of developmental expression of the human α- and β-globin genes with the top bar indicating the changes in erythropoietic tissues during development. ψ indicates an inactive pseudo gene, and the θ-gene's function is not known and is expressed at very low levels. Therefore these two genes are not indicated in the graph. Bottom: The tetrameric haemoglobins produced during the different stages of human development. Abbreviations: LCR, locus control region; HS-40, hypersensitive site -40; HbF, fetal haemoglobin; HbA1, HbA2, adult haemoglobin type 1 or type 2.

## 1.2 Eukaryotic chromatin and regulation of gene expression

Folding DNA into nucleosomal repeats and increasingly higher order chromatin structures represents an excellent way of compacting DNA. Folding is necessary to fit the DNA, which in mammals has a length of about 2 meter, into the cell nucleus with a diameter of about 10μm. This tight packaging imposes constraints on the accessibility of the DNA to diffusable regulatory proteins which play essential roles in the activation of gene transcription. There is now extensive experimental evidence implying that the regulation of the structure of chromatin contributes to the regulation of gene expression. To fully comprehend the role of chromatin in gene regulation the general structure and mechanisms regulating the dynamics of chromatin as it occurs inside the nucleus of an eukaryotic cell will be discussed first.

*Histones, nucleosome structure and chromatin organisation*

Repeating arrays of nucleosomes are the fundamental building blocks of eukaryotic chromatin (Kornberg, 1974 and 1977). The nucleosome is a complex of DNA and proteins called histones spaced at approximately 200 bp intervals along the DNA molecule. The histones form a symmetrical octamer of two of each of the H2A, H2B, H3 and H4 basic histone proteins. Around the histone core approximately 145-147 bp of the 2 nm double helix DNA is wrapped forming the nucleosome core, reducing it to form an 11 nm chromatin fiber. The coiling of the DNA creates a minor and a major groove of which the geometry plays a crucial role in the interactions of various DNA binding proteins with the DNA backbone. The core histones have amino terminal tails which pass over and between the gyres of the nucleosome wrapped DNA superhelix. These tails may make contact to neighbouring particles and are the targets of various molecular modifying events such as acetylation and phosphorylation (Richmond et al., 1984; Luger et al., 1997). The spacing between nucleosomes can vary in length and is a target for the binding of histone H1, which is required for further assembly of the nucleosomal arrays into a 30 nm fiber compacting the linear DNA 30 to 40 fold (Thoma et al., 1979; Graziano et al., 1994). Non-histone proteins are thought to further organise the 30 nm fiber into higher order structures giving rise to an overall compaction of over a thousand fold in euchromatic DNA and over ten thousand fold in heterochromatic DNA and metaphase chromosomes.

It is generally accepted that the 30 nm fiber is organised into discrete and topologically independent higher order chromatin domains or loops which are between 30-100 kb in size. These structural domains have been implied to play a role in compaction and spatial organisation of chromatin into topologically independent domains, as well as having a function in the regulation of gene expression (Gasser and Laemmli, 1986). Different extraction procedures led to the discovery of two related nuclear protein structures termed nuclear matrix and nuclear scaffold to which DNA sequences termed matrix associated regions (MARs) and scaffold attachment regions (SARs) remain bound, forming the base of DNA loops in interphase and mitotic chromosomes, respectively (Mirkovitch et al., 1984; Cockerill and Garrard, 1986a). Recently, it has been shown that many MARs co-localise with previously mapped SARs on a 500 kb region of *Drosophila* X chromosome (Iarovaia et al., 1996). Additional SARs were found to be located in the non-attached loops, suggesting a higher degree in DNA-matrix attachment during mitosis than during interphase. This reflects the rate of compaction of the chromatin fiber. Although no defined consensus sequences have been identified, S/MARs are rich in A+T base pairs (70%) and often contain topoisomerase II cleavage sites (Cockerill and Garard, 1986, Gasser et al., 1986b). S/MARs vary in size between several hundreds of base pairs and several kilobases (kb) and have been identified in many different eukaryotic species such as *Drosophila*, chicken, mouse and human (Mirkovitch et al., 1984; Cockerill and Garrard, 1986; Jarmann and Higgs, 1988; Loc et al.,

1988). S/MARs are thought to be anchored with high affinity to the nuclear scaffold through co-operative binding of matrix associated non-histone proteins. Topoisomerase II, initially identified as SC1, has been associated with the nuclear matrix and its association with S/MARs is thought to be essential for chromosome condensation and to play a role in relieving torsal stress in the DNA fiber induced during transcription (Earnshaw et al., 1985; Warburton and Earnshaw, 1997). The function of S/MAR regions in gene regulation still remains largely an enigma and few experimental clues as to the possible functions of these DNA elements have been provided. S/MARs have been frequently found in close association with regulatory sequences as enhancer elements and are able to stimulate expression of heterologous reporter genes when integrated into the genome (Gasser and Laemmli, 1986; Cockerill and Garrard, 1986; Jarmann and Higgs, 1988; reviewed by Laemmli et al., 1992). S/MARs can also act as insulator sequences, isolating domains from surrounding chromatin to form separately regulated loci. Interestingly, SARs can function as histone H1 nucleation sites thus being a site were chromatin configuration can be regulated (Zhao et al., 1993). These results suggests dual roles of S/MARs as structural components of chromatin, and as regulatory elements involved in chromatin dynamics and gene regulation, implying an intimate direct relation between chromatin structure its dynamics and gene regulation (reviewed in Laemmli et al., 1992; Strick and Laemmli, 1995).

*Chromatin structure and gene transcription*

It is generally accepted that higher order chromatin structures enforce a general negative effect on the transcriptional activity of a gene (Grunstein, 1990; Paranjape et al., 1994). At the lowest level of chromatin organisation there is both *in vitro* and *in vivo* evidence that nucleosomes function as obstacles for transcription factors trying to get to their target sequence with the purpose to activate transcription (reviewed in Felsenfeld, 1992; Svaren and Hörz, 1996). However nucleosomal arrays are present on DNA that is being transcribed or replicated, suggesting that nucleosomes do not form an insurmountable obstacle for these processes (Studitsky et al., 1995; Felsenfeld, 1996). In addition, there are also examples where the nucleosome actually plays a positive role in gene activation.

At promoter regions, positioned nucleosomes play a role in the accessibility of transcription factors to their recognition sites regulating the induction of transcription. The role of chromatin structure and histone function on transcription has been most extensively addressed in yeast. The yeast PHO5 gene is kept in a silenced state by the positioning of 4 nucleosomes in the promoter region. It becomes induced upon phosphate starvation. This process critically depends on the remodelling of the positioned nucleosomes (measured as increased accessibility to DNaseI) and is initiated by the binding of the transcriptional activator Pho4 to a low affinity-binding site in linker sequences between two positioned nucleosomes. Although the precise mechanism remains obscure the evidence suggest that all four nucleosomes are disrupted in an all or nothing fashion prior to or concomitant with the activation of gene transcription (Almer et al., 1986; Svaren and Hörz, 1997). Similar studies on interaction of the mammalian steroid glucocorticoid receptor (GR) with positioned nucleosomes in the mouse mammary tumor virus (MMTV) promoter showed how binding of the hormone activated GR to glucocorticoid response element (GRE) disrupts the nucleosome containing GRE elements thereby facilitating the binding of other factors like NF-1. NF-1 together with hormone activated GR can synergistically activate a reporter gene from a MMTV promoter although NF-1 alone showed no effect at all (Chavez et al., 1995). *In vitro* studies have shown that chromatin rearrangements can further be induced by many diverse types of transcription factors such as the yeast Gal4, Sp1, USF, TBP, TFIIIA. These transcription factors are thought to be in dynamic competition with histones for access to binding sequences in nucleosomes both on promoter and enhancer regions (Elgin, 1988;

Gross and Garrard, 1988; Steger and Workman, 1996; Felsenfeld, 1992). Nucleosome disruption by dynamic competition of transactivators does not necessarily require DNA replication although there have been reports that indicate that replication facilitates the binding of transcription factors. During activation of a promoter or enhancer region, activator proteins are thought to bind in a co-operative manner. The initial binding of a protein to relatively accessible DNA regions, which lie mostly outside or at the edge of a nucleosome, can facilitate the binding of various factors to neighbouring DNA regions, which are relatively inaccessible. The co-operative nature of binding of multiple factors to their recognition sequences stabilises the transcription-factor-protein DNA complex. This would explain why enhancer and promoter regions contain multiple factor binding sites. Binding of multiple factors to their recognition sequences additively or co-operatively increases the probability of an all-or-none formation of a stable active complex (Boyes and Felsenfeld, 1996). The exact nature of nucleosome disruption has not been investigated in most cases and can either represent a partial or complete loss of histones, histone modifications such as acetylation or phophorylation or some other allosteric changes in histone-DNA structure leading to factor accessibility and nuclease hypersensitivity (Bresnick et al., 1992; Hayes and Wolffe, 1992).

The precise positioning of nucleosomes can also bring DNA regulatory elements into closer proximity and thereby facilitate transcription. In the *Drosophila melanogaster* heat shock protein hsp26 gene promoter, the incorporation of the DNA into nucleosomal structures brings two heat shock transcription factor (HSTF) binding sites located within the linker regions into close proximity and thereby may facilitate transcription (Thomas and Elgin, 1988; Taylor et al., 1991). In a similar situation, nucleosome positioning between the promoter and enhancer of the *Xenopus laevis* vitellogenin gene brings the binding sites for the oestrogen receptor and the NF1 transcription factor into close proximity, enhancing transcription up to ten fold *in vitro* (Jackson and Benyajati, 1993; Schild et al., 1993).

Transcriptional activators are not always able to compete with histone cores for target site accessibility of the chromatin template in promoter and enhancer regions. This might be due to the relative inaccessability of the target site in the nucleosome or it might concern a relatively low affinity binding site incapable of competing with histone cores. Two distinct types of enzymatic activities can facilitate the binding of transcription factors to nucleosomal DNA, namely histone acetyltransferases and ATP-dependent chromatin remodelling.

*Histone acetylation*

Proteins containing histone acetyltransferase (HAT) activity can acetylate the positively charged lysine residues in the N-terminal tails of histones, thereby affecting their charge and function. Several lines of evidence suggest that acetylation of core histone amino termini can modify nucleosomes by decreasing the affinity of histones for DNA and destabilising the relatively inaccessible 30nm structure. This type of chromatin remodelling is thought to enhance transcription factor binding to nucleosomal DNA (reviewed in Grunstein, 1997; Luger et al., 1997; Lee et al., 1993). An association between all acetylated forms of H4 and the potentially active euchromatin was shown (Clarke et al., 1993; O'Neill et al., 1995). Hypoacetylation of lysine residues in the N-terminal tails of histone H4 is associated with X chromosome inactivation (Jeppesen and Turner, 1993). Hyperacetylation of histones has also been associated with a transcriptionally active β-globin locus exclusively in erythroid cells (Hebbes et al., 1997). These experiments provide a direct link between the acetylation status of histones and transcriptional activity. The latter study showed that acetylation of the globin genes remained associated with transcriptionally silenced genes in an active globin locus suggesting acetylation as an indicator of potential gene activity rather than transcriptional activity itself.

Euchromatin, the chromatin which decondenses during interphase and contains the actively transcribed regions of the genome, is associated with the presence of hyperacetylated histones, while the more condensed inactive heterochromatin is associated with hypoacetylated histones. Acetyltransferases and deacetylases are thought to co-operate in producing these patterns of histone modifications throughout the genome, marking these regions as either potentially transcriptionally active regions or repressed regions (Wade et al., 1997). The mechanisms underlying targeting of specific acetylation patterns to particular genes are not well understood. It is possible that transcription factors that can activate or repress transcription may associate with deacetylases or acetylases and guide these modifying enzymes to promoter and enhancer regions. On the other hand transcription factors themselves may have intrinsic HAT activity. For both cases proof has been found. The yeast transcriptional co-activator Gcn5p is required for the activation of several genes in yeast and contains intrinsic HAT activity (Kuo et al., 1996). Gcn5p is thought to function in a complex with ADA2 and ADA3 to connect activators to the basal transcription machinery thereby activating gene transcription. Targeting Gcn5p (thus HAT activity) to promoter and enhancer regions might help to remodel nucleosomal structures over these regions thereby promoting the binding of the basal transcription machinery. Human homologous of the yeast GCN5 gene have been identified and named hGCN5 and pCAF (p300/CBP associated factor). pCAF has intrinsic acetylation activity and competes with the adenoviral E1A protein for binding to p300/CBP (Yang et al., 1996). p300/CBP is for example essential as a cofactor mediating nuclear receptor-activated gene transcription in response to nuclear hormone receptor signalling (Kamei et al., 1996; Onate et al., 1995; Chakravarti et al., 1996). p300/CBP can bind to various other sequence specific factors involved in cell growth and differentiation, including CREB, c-Jun, c-Fos and c-Myb. Binding of pCAF to p300/CBP could stimulate the activation function of all these factors through enhancer and/or promoter specific histone acetylation. In contrast, the human deacetylation protein HDAC2, homologue of yeast RPD3, is associated with the transcriptional activator/repressor YY1 (Yang et al., 1996). It has been suggested that YY1 represses gene activation by tethering the histone deacetylase.

Human TAF250, a homologue of yeast TAF130 and *Drosophila* TAF230, is a basal transcription factor shown to have intrinsic HAT activity (Mizzen et al., 1996). hTAFII250 is part of a multimeric complex composed of TATA box-binding protein (TBP) and many other TBP-associated factors (TAFIIs). TAFs have been shown to provide interaction sites for distinct activators and transcription initiation factors required for activated transcription. Some TAFs, dTAFII42 and dTAFII62, have been shown to contain canonical histone folds homologous to those found in the C-terminal domains of H3 and H4, suggesting that TAFII complexes can wrap DNA of promoter regions in a nucleosomal-like structure (Wolffe and Pruss, 1996). Acetylation of nucleosomes by TAFII250 may facilitate nucleosome displacement thereby facilitating the exchange of histone and histone fold containing TAFIIs proteins.

*ATP dependent chromatin remodelling*

Various complexes have been identified to facilitate transcription by their ability to remodel or disrupt nucleosome assembly, helping transcription factors to bind to their recognition sites. The yeast SWI/SNF complex is able to change the nucleosome structure in an ATP dependent manner thereby facilitating the binding of transcription factors to the surface of a nucleosome (Petersen, 1996; Varga-Weisz and Becker, 1995). The SWI/SNF complex has a molecular weight of 2MDa and contains 11 proteins. Most of the SWI/SNF protein complexes have been shown to be components of the SRB complex, which in turn is associated to the C-terminal domain (CTD) repeats of the RNA polymerase II (pol.II)

holocomplex (Wilson et al., 1996). SWI/SNF and RNA pol.II are present in stoichiometric amounts in yeast cells, in about 2000-4000 copies per cell, leaving an estimated 100 "free" SWI/SNF copies per nucleus. The hypothesis is that the SWI/SNF proteins provide the holoenzyme with the capacity to disrupt nucleosomal DNA. It may thereby stabilise and facilitate the binding of various components of the preinitiation complex in promoter regions. The RNA pol.II holoenzyme has similar ATP-dependent chromatin disruption activity as was shown for the SWI/SNF complex, and has been shown to facilitate the binding of TBP and presumably TFIID to nucleosomal TATA sequences in promoters (Wilson et al., 1996). One of the SWI/SNF protein components is SNF2 which is a DNA stimulated ATPase essential for the remodelling activity of the complex. Mutations in the ATPase domain of SNF2 interfere with the remodelling activity *in vivo*. SWI/SNF components and their function seem to be well-conserved in eukaryotes. In human, two SNF2 homologues, hBrm and BRG-1, have been identified which are part of a SWI/SNF-complex which has been shown to disrupt chromatin structure and facilitate transcription factor binding to nucleosome associated DNA *in vitro* (Côté et al., 1994; Imbalzano et al., 1994; Kwon et al., 1994).

In *Drosophila* the SNF2 homologue Imitation-of-Switch (ISWI) has been shown to be associated with a nucleosome remodelling factor (NURF) activity observed in a cell-free embryo extract (Tsukyyama et al., 1995). NURF-activity is ATP dependent and able to disrupt nucleosomes in the hsp70 promoter in the presence of GAGA factor (Tsukiyama and Wu, 1995 and 1997). NURF is different from SWI/SNF as it consists of only four protein components and unlike SWI/SNF is present in higher levels, at an estimated 100.000 copies per cell nucleus. NURF does not associate with the RNA polymerase II holoenzyme as SWI/SNF does. In addition, the ATPase activity of the NURF-component ISWI, unlike SNF2, is stimulated by chromatin rather than DNA. The different characteristics and higher availability of NURF complexes suggest a slightly different and more general role for NURF in chromatin remodelling. More recently two remodelling complexes chromatin-accessibility complex (CHRAC) in *Drosophila* and remodelling the structure of chromatin (RSC) in *S. cerevisiae* have been identified by their ability to render chromatin accessible to binding factors. CHRAC is further able to function in chromatin assembly and to organise nucleosomes in regular arrays in an ATP dependent manner (Varga-Weisz et al., 1997). RSC has been found on the basis of homology to components in the SWI/SNF complex but is 10 fold more abundant, not associated with RNA polymerase II and unlike SWI/SNF, essential for mitotic growth. This suggests a more general role in chromatin remodelling (Cairns et al., 1996). Although the complexes mentioned here differ in many respects, the common theme in all these remodelling complexes found thus far is that perturbation of histone-DNA contacts manifest themselves by increased mobility of nucleosomes and enhanced accessibility of DNA to sequence specific DNA binding factors (Krude and Elgin, 1996). While studying these different complexes and their activities it will be interesting to see what relation there is between these complexes. Interestingly CHRAC and NURF share the ATPase containing protein ISWI, indicating that similar remodelling activities are shared by complexes with different specificity.

*DNaseI sensitivity and hypersensitive sites*

Distortions of chromatin structure of active sites can often be detected by an increase in sensitivity to nucleases such as the endonuclease DNaseI, which is able to nick individual strands of the DNA duplex in a manner that is relatively independent of the sequence. Susceptibility to DNaseI may therefore be used as an assay to measure the general accessibility of DNA regions in chromatin. An increase in DNaseI sensitivity was first noted for the globin and ovalbumin genes in expressing tissues where sensitivity was specific for active gene loci, and has now been demonstrated for a wide variety of tissue specific and

ubiquitously expressed genes (Garel and Axel, 1976; Weintraub and Groudine, 1976; Stalder et al., 1980). Sensitivity can extend over the entire region of a transcribed gene including regions up- and downstream of the genes. This type of general DNaseI sensitivity has been attributed to more common chromatin changes caused through the absence of histone H1, the acetylation of histones or undermethylation of DNA sequences (Vidali et al., 1978; Huang and Cole, 1984; Karpov et al., 1984; Schlissel and Brown, 1984; Tazi and Bird, 1990).

In addition to sensitive regions we distinguish DNaseI hypersensitive regions. These regions are particularly susceptible to digestion with DNaseI and have been identified in promoter regions as well as in distant regulatory sequences up- or downstream of many genes (Tuan et al., 1985; Forrester et al., 1986; Groudine et al., 1983). The presence of such DNaseI hypersensitive sites can be constitutive, developmentally regulated or tissue specific, and these sites are thought to represent small regions of the DNA where nucleosomes have either been disrupted or displaced, which are bound by nuclear factors (McGhee et al., 1981; Karpov et al., 1984; Solomon and Varshavsky, 1985; Felsenfeld, 1992). Constitutive sites are usually found in the promoter regions of genes poised for transcriptional activation and housekeeping genes, whereas inducible sites are most often associated with concomitant transcriptional activation (Stalder et al., 1980; Tuan et al., 1985; Forrester et al., 1986; Gross and Garrard, 1988). Both general DNaseI sensitivity and hypersensitivity can be evident prior to the onset of gene transcription and can persist after transcription has ceased. Thus at least some of the sites seem to reflect the ability of a gene or genes to be transcribed rather than the transcriptional activity itself (van Assendelft et al., 1989; Jimenez et al., 1992).

## 1.3 Regulation of eukaryotic gene expression

Although all steps involved in expressing a gene can in principle be regulated, for many genes the initiation of gene transcription is the most important point of control.

*Regulation of eukaryotic gene transcription*

Transcription is mediated through the interaction of defined proteins (transcription factors) with specific regulatory DNA sequences flanking a gene setting up the appropriate conditions for the RNA polymerase to initiate transcription. Three different RNA polymerases are involved in eukaryotic gene transcription. RNA polymerase I transcribes the genes encoding the 28S, 18S and the 5.8S ribosomal RNAs. RNA polymerase II transcribes all genes capable of encoding a protein, as well as the genes encoding the small nuclear RNAs involved in RNA splicing (classII genes). RNA polymerase III transcribes the genes encoding the transfer RNAs as well as the 5S ribosomal RNA. Since the scope of this thesis is to obtain a better understanding of the transcriptional regulation of protein encoding genes, a short and simple view of the complex picture of RNA polymerase II transcription that has emerged will be discussed.

Several DNA elements that are required in *cis* to the coding sequence have been functionally defined as elements essential for proper expression of a gene. Promoter sequences are required for proper initiation of basal levels of transcription and mark the transcriptional start site of a gene. In addition, there are sequences located at a distance from the start site of transcription that regulate the activity of promoters. These elements are known as silencers or enhancers, and repress or activate promoter mediated transcription, respectively (Ptashne, 1986). Enhancers are often part of a locus control region which profoundly influence transcription and chromatin structure of entire gene loci. An LCR element was for the first time discovered in the human β-globin locus, where the LCR is

essential in establishing an open chromatin structure in the erythroid lineage and providing high levels of expression to *cis*-linked β-globin genes (Grosveld et al., 1987). Subsequently, a number of LCRs have been found flanking many eukaryotic genes.

The multiplicity of factors and DNA control elements involved in the transcription process allows virtually unlimited possibilities to achieve accuracy in the spatial and temporal activation of gene transcription. The basal transcription machinery and the common regulatory sequences flanking eukaryotic protein encoding genes to which nuclear regulatory factors can bind, will be discussed.

*The basal transcription machinery and initiation of gene transcription*

*In vitro* studies show that the building up of the basal transcription machinery starts with the orchestrated assembly of a large preinitiation complex (PIC) at core promoters (Buratowski, 1994; Roeder, 1996). The PIC contains a well-studied collection of basal or general transcription factors (GTF) including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH and RNA polymerase II holoenzyme. The general transcription factor TFIID consists of a TATA-binding protein and eight different TBP associated factors (TAFs). TFIID is the first PIC component to assemble on the promoter through direct recognition and binding of its subcomponent TBP to the TATA box frequently found in core promoters. Binding of TFIID to the core promoter is followed by binding of TFIIA, TFIIB, TFIIF and RNA polymerase II and finally TFIIH. There is evidence that more GTFs are involved in promoter recognition. It has been shown that TAF150 (subcomponent of TFIID) can interact with a promoter region overlapping the initiator element and that TFIIA can bind to elements just upstream of the TATA box (Verrijzer et al., 1994). The binding of all three factors is thought to contribute to promoter strength and selectivity (Verrijzer et al., 1995). GTF are also the target of various activator factors bound to upstream promoter elements or enhancers which mediate their activation response through contact with specific subcomponents of the basal transcription machinery (Pugh, 1996). Multiple activators can synergise through binding of different components of the PIC thereby stabilising their interaction to promoter elements. Initiation of transcription starts with melting of promoter sequences and the release or clearance of RNA-polymerase II from the promoter bound PIC.

*Promoters*

Eukaryotic promoters of protein encoding genes are on average 100 to 200 bp in size, located immediately upstream of the transcriptional initiation site, and contain short sequences to which transcription factors can bind with the purpose to stabilise and position the transcriptional apparatus (Benoist and Chambon, 1981; Serfling et al., 1985; McKnight and Tjian, 1986; Maniatis et al., 1987). Many core promoters contain an AT-rich sequence located between 22bp to 30bp upstream of the initiation site, which is dubbed the TATA box (Breathnach and Chambon, 1981). The TATA box positions the basal transcription machinery and determines the precise site of transcription initiation. Deletions of sequences that lie between the TATA box and the normal start site of transcription give rise to new initiation sites (Grosschedl and Birnstiel, 1980a and 1980b; Benoist and Chambon, 1981; Mathis and Chambon, 1981; Kovacs and Butterworth, 1986a and 1986b). Some promoters however lack a TATA box and can have either an initiator element or have GC-rich promoter sequences. The initiator region generally overlaps the start site of transcription, has a loosely defined sequence and determines the initiation site and is capable of producing basal levels of transcription (Smale and Baltimore, 1989; Weis and Reinberg, 1992; Gill, 1994; Javahery et al., 1994). GC-rich promoters are frequently found in housekeeping genes and usually have multiple sites of transcriptional initiation.

Additional promoter elements are found upstream of the core promoter and are frequently

referred to as upstream promoter elements (UPEs). These elements can contribute to the tissue specific enhancement or repression of transcription. Binding of activator proteins to upstream promoter elements influences the assembly and stability of the promoter bound transcriptional apparatus. Stabilisation of the PIC on core promoters can facilitate multiple transcriptional initiation rounds through recurrent binding of RNA-polymerase II without having to reassemble the PIC all over again, thus increasing the expression of a gene. Some upstream promoter elements are common to many eukaryotic gene promoters such as the CCAAT box, CACC box and GC rich element frequently called SP1-box, even though these genes show very different expression patterns. A variety of ubiquitous as well as tissue and developmental stage specific expressed proteins can recognise and bind CCAAT elements and thereby contribute to either constitutive or more tissue and stage specific gene expression. Some factors that can bind to promoter elements are thought to have a more structural role in that they can bend the DNA fiber and allow factors bound to upstream sequences to approach core-promoter bound factors (Dyer et al., 1996). The binding of TBP to the TATA box has been shown to bend the DNA fiber adding to the topology and structure of promoter elements critical for the onset of transcriptional activation (Kim et al., 1993).

*Enhancers and silencers*

Enhancers are *cis*-acting elements which are functionally defined by their ability to stimulate the basal level of transcription of an adjacent promoter after the binding of particular transcription factors. Enhancers activate the level of transcription in an orientation independent manner and relatively independent of the distance from the initiation site (Banjeri et al., 1981; Moreau et al., 1981; Serfling et al., 1985; Treisman and Maniatis, 1985; Weber and Schaffner, 1985). Enhancers themselves can not initiate transcription, but have sequence motifs in common with upstream promoter elements. It is therefore not remarkable that both elements can physically and functionally overlap (Muller et al., 1988; Ho and Leiden, 1990; Lin et al., 1993). Enhancers can be found upstream, downstream, or in the body of a transcribed unit. Some enhancers flank promoter regions of more than one gene, however their function in activating gene transcription can be limited to only a single gene depending on the specificity of promoter sequences (Merli et al., 1996).

Silencer elements are similar to enhancers in that they contain a combination of sequence motifs and function in an orientation and distance independent manner. They differ from enhancers in that they have a repressing effect on transcription of genes. Silencer elements were first identified in the yeast mating type locus (Brand et al., 1985) and have now been identified in many tissue specifically expressed genes such as the $\alpha$ T-cell receptor gene, the chicken lysozyme gene and the human embryonic $\varepsilon$-globin gene (Winoto and Baltimore, 1989; Baniahmad et al., 1990; Cao et al., 1989).

In the regulation of gene expression, silencer elements co-operate with enhancer elements to produce co-ordinate tissue and developmental stage specific gene expression (Huang et al., 1993, Trepicchio et al., 1993).

*Mechanisms of enhancer and silencer function*

Different combinations of sequence motifs are important in enhancer activity, suggesting that transcription factors bound to adjacent sites interact, and that these protein-protein interactions are needed for enhancer activity (Dynan, 1989). Upon binding of a transcription factor to its recognition sequence it is thought to change the local chromatin structure via disruption or displacement of a nucleosome. Binding of multiple transcription factors to adjacent sites in an enhancer might work in synergy to displace the nucleosomes and stabilise their binding. Once regulatory proteins are bound to the enhancer sequence they are thought to influence the assembly or stability of the promoter bound transcriptional apparatus. Several

mechanisms for enhancer function have been proposed. Prevailing models suggest that enhancer bound protein complexes interact directly via protein-protein contact to the promoter bound factors via looping out of intervening DNA sequences. Alternative models predict that enhancers only increase the probability of forming a stable transcription complex at the promoter by antagonising repressive chromatin structures or can function as nucleation site of a transcriptional activity which scans along the DNA fiber (Weintraub, 1988; Walters et al., 1995 and 1996; Tuan et al., 1992; Kong et al., 1997). Enhancers have also been shown to act in *trans* (one allele acting on the other allele) to stimulate transcription, a phenomenon called transvection. Enhancers have further been shown to stimulate transcription of a β-globin promoter when linked via a biotin-streptavidin bridge (Müller et al., 1989). Both findings can most easily be explained with the looping model rather than the alternatives. Nevertheless, the exact mechanism involved in enhancer stimulated transcription is still a topic of debate. These mechanisms will be more elaborately discussed later in the context of how the LCR can influence transcriptional activity of the individual β-globin genes.

How silencers repress gene transcription is not known. One can envision several ways in which silencers can block the activation function of enhancers. Silencer elements may compete for promoter interaction rendering it silent. Interestingly silencers are frequently located in closer proximity to promoter regions than enhancers, giving silencers an advantage to interact with the promoter. In analogy, silencers may sequester enhancer elements from promoter interaction. Silencers may also distort local chromatin structures thereby affecting the function of neighbouring regulatory sequences.

*Locus control regions (LCRs)*

LCRs are defined as regulatory elements that define active domains of gene expression in a dominant fashion and are required for the proper regulation of tissue specific expression of many vertebrate genes. LCRs resemble classical enhancers in that they contain numerous binding sites for transcription factors and can enhance transcription from large distances, either up- or downstream of genes and independent of orientation of the transcriptional start site (Talbot et al., 1989; Dillon and Grosveld, 1993; Kioussis and Festenstein, 1997). However, LCRs are different from enhancers in that they govern gene expression independent of the site of integration in stable cell transfections and transgenic animals (Grosveld et al., 1987; Blom et al., 1989). Classical enhancers, on the other hand, are influenced in their activity by the site of chromosomal integration presumably due to the effect of the surrounding chromatin in which they have integrated (Wilson et al., 1990). LCRs therefore seem to be able to override negative chromatin effects. The first LCR identified was the human β-globin LCR, which provides high levels of erythroid specific expression to a *cis*-linked adult β-globin in a copy number dependent, integration site independent manner (Grosveld et al., 1987). The β-globin LCR consists of five regions of about 200 to 300 bp in size, which were identified as erythroid specific, developmentally stable DNaseI hypersensitive sites (HS), HS1, 2, 3, 4 and 5 (see Figure 3 and 4). It has been shown that different HS-sites might provide the LCR with distinct functions as enhancer activity predominates in HS2 and chromatin opening activity in HS3. LCR activity on the chicken lysozyme gene is provided when the entire gene locus carrying the full set of regulatory elements including three enhancers, one silencer and a complex promoter, is used as a construct to make transgenic animals (approximately 20 kb). Deletion of any single HS-site from the β-globin LCR as well as deletion of any one enhancer region in the chicken lysozyme locus abolishes position independent expression, suggesting a concept in which all HS-sites or enhancers act as a single entity (Milot et al., 1996; Bonifer et al., 1994). LCRs have now been identified for many other vertebrate genes which are abundantly expressed in a tissue specific manner like the rat whey acidic protein gene, human growth hormone, human

CD2, chicken lysosyme, mouse methalothionein, mouse T-cell receptor and the human adenosine deaminase (Dale et al., 1992; Jones et al., 1995; Greaves et al., 1989; Bonifer et al., 1990; Palmitter et al., 1993).

## 1.4 Transcriptional regulation of β-globin gene expression

In order to characterise DNA regions involved in the regulation of human β-globin genes, expression studies have been performed using different assay systems such as patient material, transgenic mice, and tissue culture cell lines (discussed below). From these studies it is clear that the β-globin locus is regulated at different levels. Introduction of individual gene constructs including proximal regulatory elements in transgenic mice provided erythroid specific expression with a similar developmental stage specific pattern as their endogenous mouse counterparts (Magram et al., 1985; Chada et al., 1985; Townes et al., 1985; Kollias et al., 1986). Thus the *cis*-regulatory elements that can confer the erythroid and stage specific expression appear to be closely associated with the genes. However, additional sequences were missing, since the expression of these constructs was very low compared to the mouse endogenous globin genes and was independent of copy number. This subsequently led to the discovery of the β-globin locus control region at the 5' end of the gene cluster, which regulates the entire locus. Thus, the regulation of β-globin gene transcription occurs through interaction of erythroid specific and ubiquitous *trans*-acting proteins with *cis*-regulatory sequences surrounding the gene, that include promoters, enhancers, silencers, and the more gene distal regulatory elements of the locus control region (LCR) (Grosveld et al., 1993).

*Approaches in the study of human β-globin gene regulation*

Natural models for the study of tissue specific transcriptional regulation have been provided by a large number of genetic defects involving the human β-globin locus that are characterised by altered or reduced synthesis of the different β-like globin chains and which are responsible for a heterogeneous group of genetic diseases known as HPFH and β-thalassaemias (Collins and Weissman, 1984; Orkin, 1986; Stamatoyannopoulos and Nienhuis, 1987). Alterations in the genetic constitution of patients can vary from very subtle single base pair substitution to large genomic deletions of up to 100kb or more as in the case of Dutch and English deletion type thalassaemias. Additional insights into the mechanism of globin gene switching have been acquired by the studies of mutations that increase the production of foetal haemoglobin in adult blood (Wood, 1993; Rochette et al., 1994).

Regulation of β-globin gene expression has been extensively studied using different transformed human and mouse erythroid cell lines. The Friend virus transformed Murine erythroleukemia (MEL) cell line is frequently used and represents a mouse erythroid precursor cell blocked at the pre-erythroblast stage of its differentiation (e.g. Antoniou et al., 1991). These cells can be chemically induced to terminal differentiation when cultured *in vitro* much like normal red-cell maturation. During induced differentiation MEL cells undergo two rounds of cell division and show accumulation of the mouse endogenous adult globin chains (α; βmaj and βmin; Marks and Rifkind, 1978), increase in heme biosynthesis, chromatin condensation and other morphological changes culminating in a cessation of cell division and ultimately anucleation (Volloch and Housman, 1982; Hofer et al., 1982). MEL cells therefore provide an excellent model system for studying erythroid specific regulation of gene expression associated with differentiation. Striking similarities are seen between results obtained using MEL cells to those obtained *in vivo* using mice (Fraser and Curtis, 1987). Human cell lines that are commonly used are K562 and human erythroleukemia (HEL),

isolated form a chronic myelogenous leukemia and an erythroleukemia, respectively. Both are of an erythroid/megakaryocytic cell type that can be chemically induced *in vitro*, though unlike MEL cells they cannot go through complete terminal differentiation. Both cell lines stably express the embryonic and foetal globin genes (Martin and Papayannopoulou, 1982). The use of cultured cell lines has its limitations because they are not normal cells due to their neoplastic nature. A major disadvantage of using cell lines is that studies on developmental aspects of globin gene expression can not be performed.

The most valuable approach in the study of globin gene switching mechanisms has been the development of methods to introduce foreign DNA into the germ line of mice (Gordon et al., 1980; Constantini and Lacy, 1981). Foreign DNA is introduced into the germ-line of mice via microinjection of femtogram amounts ($10^{-15}$gram) of DNA into one of the pro-nuclei of fertilised mouse eggs. Usually injected DNA integrates at a single random site in the mouse genome and as a multicopy head-to-tail tandem repeat which is inherited in a Mendelian fashion (Gordon et al., 1980; Stewart et al., 1982). The foreign DNA is thought to integrate at random with no additional selection for its expression, as is the case in making transgenic cell lines. In addition, the use of transgenic mice allows for temporal and tissue specific studies of gene expression during development *in vivo*.

*Gene proximal control elements*

The proximal or minimal β-like globin gene promoter contains three conserved *cis*-regulatory sequences which are all critical for transcriptional activation in both erythroid and non-erythroid cells including a TATA box around -30, a CCAAT box between -70 and -90 and a CACC box lying even further upstream at -90 and -110 (Grosveld et al., 1982; Dierks et al., 1983; Myers et al., 1986). The TATA box in all β-globin genes positions the basal transcription machinery through the binding of the multiprotein basal transcription factor TFIID (Antoniou et al., 1995). In the case of the adult chicken β-globin gene it has been reported that the TATA box can bind the erythroid specific factor GATA-1 (Fong and Emerson, 1992). In contrast transfection experiments in MEL cells of DNA constructs containing the human β-globin gene promoter indicated no functional significance of GATA-1 binding to the TATA box (Antoniou et al., 1995). Induction of tissue specific regulation of transcription is mediated through the CCAAT and the CACC boxes in the minimal promoter (Antoniou and Grosveld, 1990). In the case of the human δ-globin gene mutations in the CCAAT and CACC box sequences are thought to be the cause of the low level of δ-globin gene expression. Correction of these two elements into sequences mimicking the sites of the β-globin promoter restored high levels, erythroid specific expression of the δ-globin gene (Donze et al., 1996; Tang et al., 1997). The γ-globin gene minimal promoters differ from the β-globin gene minimal promoter in that the γ-globin genes have a single CACC box upstream of two duplicated CCAAT boxes centred on positions -113 and –86. Point mutations that map to these CCAAT elements have been correlated with HPFH conditions suggesting a role of these elements in the transcriptional silencing of the γ-genes during development (Wood, 1993). The β-globin gene-promoter contains a single CCAAT box at –75 and duplicated CACC boxes centred on positions -105 and -90. Only the most proximal β-globin CACC box appears to be necessary for maximum expression and several point mutations have been documented which result in decreased expression of the β-gene in foetal and adult erythrocytes as well as in tissue culture cells (Orkin, 1984 and 1988; Treisman et al., 1983; Feng et al., 1994).

Several ubiquitously and tissue specifically expressed factors have been identified which can bind the CCAAT box in the β-globin promoters with varying affinity namely CP1, CP2, NF1, NF-E6, GATA-1 and CCAAT displacement factor (CDF) (Chodosh et al., 1988; deBoer

et al., 1988; Superti-Furga et al., 1988; Mantovani et al., 1989; Berry et al., 1992). NF-E3 can bind to the sequence immediately flanking the distal Aγ-globin CCAAT box which is abolished in the case of the HPFH condition caused by a base pair substitution at the -117 position as well as the 13 bp deletion of the distal CCAAT box region (reviewed in Wood, 1993; Ronchi et al., 1996; Superti-Furga et al., 1988; Mantovani et al., 1989). This suggests that NF-E3 may function as a repressor protein preventing binding of other CCAAT binding proteins. The CACC box can be recognised by several ubiquitously as well as tissue specifically expressed nuclear factors. These include SP1 and TEF2/BKLF but in the case of the adult β-globin proximal CACC box only the erythroid specific factor EKLF seems to be essential (Xiao et al., 1987; Miller and Bieker, 1993; Crossley et al., 1996; Perkins et al., 1996; Wijgerde et al., 1996 or Chapter 3 of this thesis).

Two less well-defined, though evolutionarily conserved, gene specific minimal promoter elements have been identified in both the γ- and β-globin minimal promoters. In the case of the β-gene a direct repeat element (β-DRE) is located between the CCAAT and TATA boxes. The β-DRE consists of a imperfect direct repeat of a 10 bp motif to which a factor called βDRf is thought to bind thereby introducing a bend in the DNA helix (Stuve and Myers, 1990 and 1993). The γ-globin promoter has been reported to contain a so called stage selector element (SSE), first identified in the chicken β-globin promoter, which is centred at the -50 position and held responsible for preferred foetal expression of the γ-genes (Choi and Engel, 1988; Jane et al., 1992). The SSE is recognised and bound by the stage selector protein (SSP), an erythroid foetal specific protein complex with CP2 as a subcomponent (Jane et al., 1992 and 1995). However analysis of transgenic mice in which the role of the SSE was tested through simple deletion revealed little to no effect on the expression of the foetal γ-genes (Grosveld et al, unpublished). Nevertheless it is thought that some of these differences in the minimal promoter regions of the γ- and β-genes are, at least in part, responsible for developmental specific transcriptional regulation.

The recognition sequence (T/A)GATA(A/G) is found in all β-globin promoters and upstream promoter sequences and is recognised by the lineage restricted transcription factor GATA-1 (previously called NF-E1, Eryf-1 or GF1). In addition potential binding sites for a wide variety of transcription factors have been identified in the region located upstream of the minimal promoter, many of which are important for tissue specific inducible expression of the β-globin genes (deBoer et al., 1988; Antoniou, 1988; Yu et al., 1990; Gong et al., 1991; Gumucio et al., 1991; Trepicchio et al., 1993).

The human γ-globin genes have a single enhancer, which is located approximately 400bp downstream of the polyadenylation signal of the Aγ-gene and comprises about 700bp. The enhancer is characterised by two erythroid specific DNaseI hypersensitive sites and in transient transfections the enhancer can activate reporter genes to various levels (6-23 fold) in both erythroid and non-erythroid cell lines (Bodine and Ley, 1987). DNaseI protection and gel mobility shift assay revealed 8 footprints, of which three are binding sites for the erythroid specific activator GATA-1 and two are specific for the promoter enhancer factor-1 (PEF-1). The enhancer has been suggested to act as a silencer since transgenic mice lacking this element inappropriately express a low level of γ-globin in adult erythroid cells (Enver et al., 1989).

Two human β-globin gene enhancers have been identified that stimulated the level of expression of linked reporter genes in transgenic erythroid cells and in MEL cells (Behringer et al., 1987; Kollias et al., 1987; Antoniou et al., 1988). One element is located around the EcoRI site within the third exon (intragenic enhancer) and the other is located downstream of the gene (3'enhancer). The 3' enhancer minimal region which has tissue- and stage-specific enhancing activity is located within a 250 bp PstI restriction fragment approximately 600 bp

downstream of the polyadenylation signal. DNaseI footprinting and mobility shift assays revealed four different nuclear protein binding regions each can be recognised by the erythroid specific transcription factor GATA-1 but also by additional non-erythroid specific factors (Wall et al., 1988). Deletion of the enhancer severely down regulated the expression of the adult β-globin gene in transgenic mice (Bungert et al., 1997).

A silencer element has been identified between –177 and –392 bp upstream of the human ε-globin gene promoter (Cao et al., 1989). Silencing is mediated by the binding of two erythroid specific GATA-1 proteins and a ubiquitous YY1 protein and deletion of the silencer element resulted in continuous expression during late foetal and adult stages (Raich et al., 1992 and 1995). However deletion of a 125bp (-304/-179) minimal silencer region containing the two GATA-1 and YY1 sites was also shown to be required for activation of the embryonic ε-globin and the foetal γ-globin genes during the early yolk sac stage of erythroid development (Bungert et al., 1997).

*The human β-globin locus control region*

Evidence for the presence of a region important in the regulation of the entire β-globin locus first became apparent from the study of a Dutch γδβ-thalassaemic patient and later also by studies on Hispanic γδβ-thalassaemia (Kioussis et al., 1983; Driscoll et al., 1989; see Figure 4). The Dutch thalassaemia contained a large deletion (~100kb) which removed sequences between 60kb upstream of the ε-globin gene to about 2.5 kb upstream of the β-globin gene (Kioussis et al., 1983; Taramelli et al., 1986; van der Ploeg et al., 1980). The β-globin gene was still intact from the allele that contained the deletion but it was not expressed in erythroid cells. Moreover it was shown that the remaining sequences of the deleted locus were hypermethylated and DNaseI insensitive, both indicative of silenced regions (Kioussis et al., 1983). Later studies on the Hispanic thalassaemia showed that deletion of the LCR affected the chromatin structure over a 200 kb region and rendered the locus from normally early to late replicating in the S-phase of the cell cycle (Forrester et al., 1990). DNaseI studies identified five major erythroid specific developmentally stable DNaseI hypersensitive sites spread over an approximately 16kb region, 6 kb upstream of the embryonic ε-globin gene (Tuan et al., 1985; Forrester, 1986). The functional significance of these DNaseI regions became apparent when they were used to drive expression of the human β-globin genes in transgenic mice (Grosveld et al., 1987). A 21 kb DNA region covering the whole LCR region is called miniLCR and was shown to drive full expression of a linked β-globin gene in transgenic mice (Figure 4). The functional activity of the LCR was retained in a 6.5kb construct, called microLCR (μLCR), in which small (1-2kb) regions containing the individual 5'HS sites were linked together, and tested in transgenic mice and stably transformed MEL cells (Talbot et al., 1989; Figure 4). An additional HS site (3'HS 1) was found some 10kb downstream of the adult β-globin gene. However, studies in transgenic mice showed no effect on LCR activity when it was absent and the Dutch and Hispanic thalassaemias in which the adult β-gene as well as the 3'HS site are still intact and present, indicated that no major functional activity resides at this site.

The contribution of the individual HS sites to β-globin gene expression was tested in transgenic mice (Fraser et al., 1990) and transient as well as stably transfected tissue culture cells (Tuan et al., 1989; Collis et al., 1990; Ney et al., 1990; Moon and Ley, 1991). The main activity of the β-globin LCR is associated with HS2 HS3 and HS4 (Forrester et al., 1989; Collis et al., 1990; Fraser et al., 1990 and 1993; Lowrey et al., 1992), which is in agreement with the data from the Hispanic γδβ-thalassaemia where an intact HS1 is still present but does not result in detectable LCR activity (Driscoll et al., 1989). HS2, HS3 and HS4 provided copy number dependent, integration independent expression to a linked human β-gene in multicopy

transgenic mice (Fraser et al., 1990). These results indicated that HS3 showed nearly 70%, HS2 and HS4 each about 30% while HS1 showed less than 10% of full LCR activity. Stable transfections in MEL cells of the single LCR HS sites more or less confirmed these results (Collis et al., 1990). In transient transfection assays only HS2 behaved as a classical enhancer (Tuan et al., 1989). HS3 was the only site able to provide high levels of erythroid specific expression to a linked globin gene when integrated as a single copy in transgenic mice suggesting that a chromatin opening activity resides in HS3 (Ellis et al., 1996). In addition, experiments in transgenic mice showed some developmental or gene specificity of the LCR HS3 and HS4 sites (Fraser et al., 1993). These experiments suggest that the different 5'HS sites have distinct functions and contribute additively to the function of the LCR in regulating globin gene expression. Analysis of globin gene expression in transgenic mice containing single LCR HS site deletions in 70kb human β-globin locus constructs were shown to be prone to position effects due to a crippled function of the LCR. The fact that all sites are required for position independence suggests that the individual HS sites have non-redundant functions during development and that the LCR acts as a single integral entity. However, expression analysis of individual HS site knockouts of the murine β-globin locus showed no significant effect on the timing of gene expression and only modest reduction of globin gene expression. Deletions of murine HS2 and HS3 resulted in mild reductions of globin gene expression in embryonic cells and a 30% reduction of adult β-globin RNA levels (*βmaj* and *βmin* together) in foetal and adult red cells whereas deletion of HS1 and HS4 showed even milder reductions in the level of globin gene expression (Hug et al., 1996; Fiering et al., 1995). These latter experiments suggest that no single HS site has a unique function in gene switching but only in the case of HS2 and HS3 are each associated with 30% of the adult β-globin gene expression in adult red blood cells. A simple explanation for the different results could be that the mouse and human loci are not completely homologous in structure and function and thus are regulated by different mechanisms. However, the human and murine LCR HS1-4 exhibit great sequence homology in their core regions (>60%) and the structure and function of both LCRs seem to be largely conserved (Moon and Ley, 1990; Hug et al., 1992; Jimenez et al., 1992). In addition natural deletion of human HS1 does not result in any alteration of adult globin gene expression mimicking the murine HS1 knockout results (Kulozik et al., 1991). A more satisfactory explanation would be that in the case of the transgenes the LCR integrates in some cases in a more challenging chromatin environment than the native surroundings of the endogenous locus. Integrations into different chromatin surroundings can result in a mild to severe reduction of the overall expression due to two different types of position effects, namely position effect variegation (PEV) and a cell timing effect (Milot et al., 1996). The cell timing effect is still under experimental scrutiny and no molecular explanation has yet been found. The silencing effect in PEV is envisioned as the spreading of heterochromatin from a heterochromatic region which functions as the nucleation site of heterochromatin protein assembly. Heterochromatin silencing and the chromatin opening and transcriptional activation activity of the human β-LCR are thought to be in a continuous competition. A full intact LCR ensures an open chromatin structure and active gene transcription in all cells at all times (Milot et al., 1996). Any HS site deletion cripples the LCR such that it looses strength in its competition with heterochromatin spreading. This is measured as a drop in the levels of gene transcription and DNaseI sensitivity which is reflected at the single cell level by a reduced number of cells with transcriptionally active genes. This phenomenon is inherited in an as yet unexplained clonal fashion (Milot et al., 1996).

In order to understand the underlying molecular mechanisms of human β-globin LCR activity, the minimal core fragments of HS2-4 with their protein binding properties have been analysed (Figure 4). The regions that retained LCR activity were determined to reside in 200-

300bp fragments of DNA. These minimal, or so called core regions, were subsequently analysed *in vitro* (Talbot et al., 1990; Philipsen et al., 1990; Pruzina et al., 1991; Lowrey et al., 1992) and *in vivo* (Reddy and Shen, 1992; Ikuta and Kan, 1991; Straus and Orkin, 1992) for the presence of erythroid specific and ubiquitous transacting factor binding sites. A summary of the protein binding sites in the core regions of HS2, HS3, and HS4 of the human β-globin LCR is illustrated in Figure 4. All sites contain similar binding sites though in different combinations. All HS core regions have closely spaced binding sites for GATA-1, several G-rich regions to which factors as Sp1, TEF2/BKLF and EKLF can bind, and AP1 or NF-E2 consensus binding sites. Strikingly, the same three factor binding motifs are found in promoter and enhancer regions of the individual globin genes. In the case of HS3 the minimal combination (indicated in Figure 4 as core) that provides position independent expression in mice is the presence of two GATA-1 sites on either side of a G-rich motif which have been shown to be occupied by proteins by *in vivo* footprinting in erythroid cells (Philipsen et al., 1993; Strauss and Orkin, 1992). The core region of HS2 contains a dimer of AP-1 sites which bind the erythroid specific nuclear factor NF-E2. The strong erythroid enhancer activity of HS2 has been shown to depend on the presence of the tandem NF-E2 sites (Ney et al., 1990a and 1990b; Caterina et al., 1994a). NF-E2 was originally identified as an erythroid specific DNA binding activity and is a heterodimeric factor that consists of a large basic leucine zipper (bZip) subunit (p45) and a small bZip subunit (MafK) (Andrews et al., 1993a and 1993b; Igarashi et al., 1994). The large subunit p45 NF-E2 belongs to the Cap'n'collar (CNC)-type bZip protein family with members such as Nrf1/LCR-F1/TCF11 and Nrf2 (Caterina et al., 1994b; Chan et al., 1993; Moi et al., 1994). However p45 is dispensable for LCR activation and globin gene transcription (Shivdasani and Orkin, 1995). Novel candidates as interaction partners for MafK which as heterodimers can bind to NF-E2 sites, have recently been found with the discovery of a novel bZip transcription factors called Bach1 and Bach2 (Oyake et al., 1996). In addition HS2 contains three sites for GATA-1, YY1 and USF *in vitro* (Talbot and Grosveld, 1991; Ellis et al., 1993). Much of the enhancer activity of HS2 resides in the NF-E2 sites though all protein binding sites are required for full HS2 activity (Caterina et al., 1991; Talbot and Grosveld, 1991; Lui et al., 1992). Except for YY1 all sites have been shown to bind factors *in vivo* (Ikuta and Kan, 1991; Reddy and Shen, 1992). The core region of HS4 maps to a 280 bp region, containing binding sites for the erythroid specific proteins GATA-1 and NF-E2 and a number of ubiquitous factors as jun/fos, Sp1 and TEF2/BKLF (Pruzina et al., 1991).

The minimal requirement through which the LCR seems to be able to drive activated levels of expression is the presence of a TATA box, a CCAAT box and a CACC box in promoter regions (Antoniou and Grosveld, 1990). The LCR has been shown to drive expression of heterologous promoters in an inducible erythroid specific manner of the herpes simplex virus thymidine kinase promoter, the murine Thy-1 promoter and a minimal hsp68 promoter which all contain functional CCAAT, CACC and TATA boxes (Talbot et al., 1989; Blom et al., 1989; Tewari et al., 1996). However the LCR does not activate expression from the murine histone H4 promoter which seem to reflect the absence of these two elements (Antoniou and Grosveld, 1990).

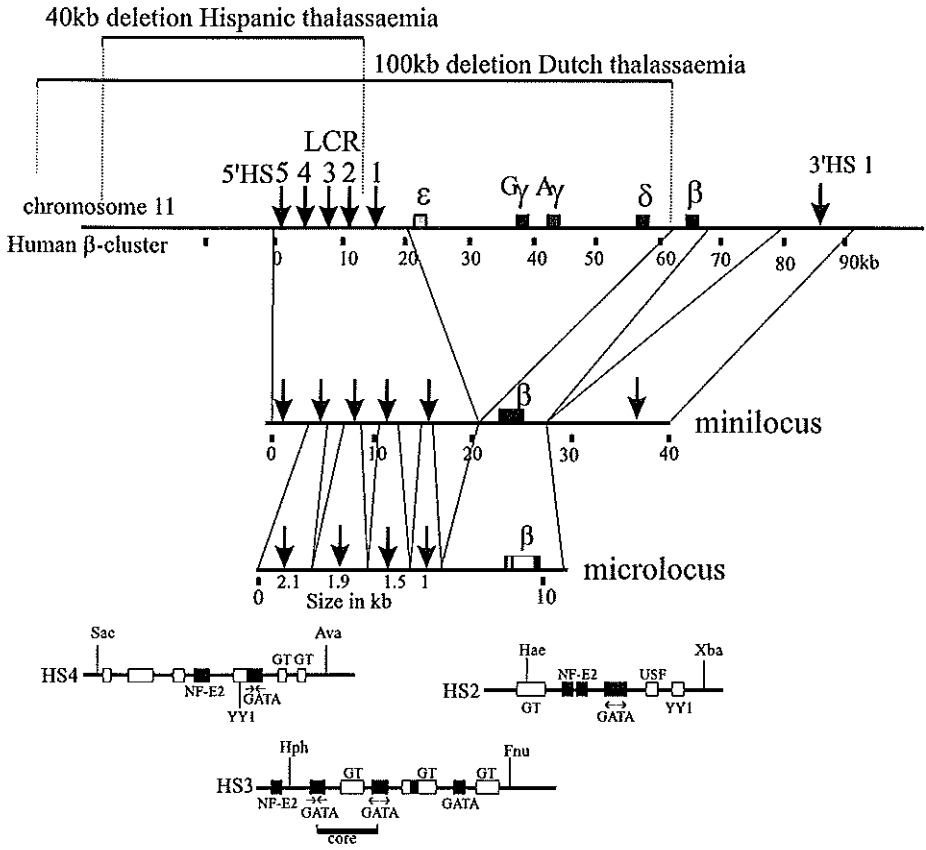



**Figure 4.** The human β-globin locus control region (LCR)
Top: The lines above the human β-globin locus indicate the deletion breakpoints of the Dutch (van der Ploeg et al., 1980) and Hispanic (Driscoll et al., 1989) γδβ-thalassaemias relative to the genes and the DNaseI hypersensitive sites (vertical arrows with number) of the LCR (Tuan et al., 1985). 3'HS site is also shown and is located more than 20 kb downstream of the adult β-gene. Middle: The β-globin minilocus consists of a 21 kb LCR fragment containing all five major 5'HS sites and a 12 kb fragment with 3'HS1 which were cloned upstream and downstream of the human adult β-globin gene, respectively (Grosveld et al., 1987). In the case of the β-globin microlocus, the individual HS sites were cloned as small 1 to 2 kb fragments upstream of the human adult β-globin gene which in transgenic mice gave similar LCR activity as the minilocus construct (Talbot et al., 1989). Bottom: Schematic summary of the protein factor binding sites in the minimal regions harbouring LCR-like activity of the HS2, HS3, and HS4 regions. Shaded boxes indicate binding sites for erythroid specific factors, black boxes indicate binding sites for ubiquitous factors. Arrows indicate the direction of the minimal DNA sequence motif recognised by GATA-1. The DNA fragment containing the minimal combination of transcription factors which is still able to provide copy number dependent expression is dubbed the core of HS3 and is indicated by a thick black line (Philipsen et al., 1992).

*Trans-acting factors in the regulation of human β-globin gene expression*

Most transcription factors contain a DNA binding domain and a separable activator or repressor domain (Mitchell and Tjian, 1989). Swapping experiments have shown that these two domains can frequently be separated from each other and still remain functional. Most transcription factors are grouped into families, which share a common structure or function based on homologies between DNA binding or protein-protein interaction domains. Although no strict correlation is seen between the function of the regulatory domain (repressor/activator) and the amino acid composition, activator domains have been frequently found to be rich in acidic, glutamine or proline amino acid residues, while repressor domains have been found to be either rich in basic amino acids or alanine rich (Han and Manley, 1993; Saha et al., 1993; Licht et al., 1994).

The fact that stage and tissue specific factors play an important role in globin gene regulation has been unambiguously shown in expression analysis in interspecific heterokaryons formed between fusions of adult specific MEL cells and embryonic/fetal specific human erythroid K562 cells, or a variety of non-erythroid cell types (Baron and Maniatis, 1986). The nuclei in transient heterokaryons do not fuse and therefore the rapid reprogramming of globin gene expression that was observed must occur in *trans* through the action of erythroid stage specific nuclear factors. This indicates that switching in the expression of the human globin genes during development is, at least in part, regulated by the action of stage and tissue specific activators and repressors.

The different globin gene promoters, enhancers, silencers and elements of the β-globin locus control region bind a strikingly limited array of transcription factor proteins often with similar and closely spaced binding sequence arrangements. A number of factors that bind to the regulatory *cis*-elements have been isolated from erythroid cells through their specific DNA sequence binding activities using various *in vitro* DNA binding assays. These factors can be divided into those that are expressed exclusively in the erythroid lineage, such as GATA-1, NF-E2, EKLF, TAL1/SCL1, RBTN2 and NF-E4, and more widely expressed transcription factors in erythroid cells, such as GATA-2, YY1, Sp1 and TEF2/BKLF (reviewed in Bungert and Engel, 1996; Baron, 1997). Numerous experiments have been carried out on the activity and function of these proteins. In this discussion, I will limit my attention on only three erythroid specific factors and one non-erythroid factor, namely GATA-1, EKLF, NF-E2 and Sp1, respectively.

GATA-1 was the first erythroid specific transcription factor to be identified and cloned and recognises the consensus sequence (T/A)GATA(A/G) found in most if not all regulatory sequences of the β-globin locus (Tsai et al., 1989; Evans and Felsenfeld, 1989). However, later experiments have shown that GATA-1 is also expressed in uncommitted haemapoietic precursor cells, megakaryocytes, eosinophils, mast cells and Sertoli cells. GATA-1 is the founding member of a family of GATA transcription factors that share homology in their DNA binding domain, which consists of two highly conserved zinc fingers. The amino acid sequences outside these finger domains have not been conserved. GATA-1 can activate transcription upon co-transfection with reporter gene constructs. GATA-1 can also synergise and associate with many other factors present in erythroid cells such as YY1, Sp1 and EKLF, and frequently binding sites for these proteins are found in close proximity. Genetic ablation of GATA-1 expression through homologous targeting in ES cells, revealed an essential role of GATA-1 during erythroid differentiation (Pevny et al., 1991). Additional experiments which make use of chimaeric mice between normal and GATA-1 targeted ES-cells showed that the arrest of erythroid differentiation takes place at the proerythroblast stage (Simon et al., 1992). Interestingly, GATA-2 is upregulated in GATA-1 homozygous knockout mice and overexpression of GATA-2 in avian erythroid cells resulted in an arrest of erythroid differentiation and a higher level of proliferation of erythroid precursor cells (Briegel et al.,

1993; Weiss et al., 1994). GATA-2 is more widely expressed. Homozygous knockout mutant mice and chimaeric (normal plus knockout ES cells) mice have implied an important role for GATA-2 in proliferation of early haematopoietic progenitor cells (Tsai et al., 1994).

Erythroid Krüppel-like factor (EKLF) was isolated after enrichment of erythroid cell specific transcripts by subtractive cloning between a mouse MEL cell line and a mouse monocyte macrophage cell line (Miller and Bieker, 1993). Later the human homologue was isolated which shows a great degree of structural and sequence homology (Bieker, 1996). EKLF contains three zinc fingers with homology to those found in the Krüppel family of transcription factors able to recognise and bind the DNA sequence CCACACCCT, an essential element in the adult β-globin promoter and related regulatory sequences (Feng et al., 1994). Subsequent co-transfection studies in K562 cells demonstrated that EKLF prefers to activate a reporter gene driven by the adult β-globin promoter over one that drives expression of the human γ-gene promoter (Donze et al., 1995). Naturally occurring point mutations that map to the β-globin CACC box have been implicated as causative of severe β-thalassaemias (Reviewed in Thein, 1993). Testing three CACC box mutations which are known to give rise to thalassaemic phenotypes in human, revealed reduced transactivation of linked reporter genes by EKLF (Feng et al., 1993). Although EKLF is expressed in primitive as well as in definitive erythroid cells (Southwood et al., 1996), targeted disruption of the EKLF gene in murine ES cells revealed only an essential role in adult β-globin gene expression (Nuez et al., 1995; Perkins et al., 1995). EKLF-/-mice are normal during embryonic stages, but die *in utero* around days 14-15 of gestation from a severe anaemia due to an almost complete lack of adult β-globin gene expression. In a search for additional CACC box binding transcription factors a new Krüppel family member was identified called BKLF (Crossley et al., 1996). BKLF was shown to correspond to a gel shift activity called TEF2, which was first described in the context of the SV40 enhancer. BKLF is more widely expressed including primitive and definitive erythroid cells and can bind with high affinity to many different CACC sequences found in erythroid promoters. This suggest that BKLF might be a good candidate protein for the regulation of other genes containing a CACC box like the murine embryonic genes.

Sp1 is a well-characterised transcription factor that has been implicated in the regulation of expression of numerous genes by *in vitro* and *in vivo* assays. Sp1 belongs to the growing family of proteins with homologous zinc finger domains that can recognise and bind GT or GC rich sequences. Several such sequences are found throughout the human β-globin locus including the core regions of the β-globin LCR HS elements. Recent analysis from our lab indicated that the globin genes are normally expressed in Sp1 homozygous knock out mice (Marin et al., 1997). This either suggests that either Sp1 is not important in the regulation of the globin genes, or another factor(s) can substitute for Sp1, with candidates EKLF, TEF2/BKLF or Sp3.

The transcription factor NF-E2 was originally identified as an erythroid specific binding activity expressed in erythroid, megakaryocytic, and mast cells and shown to bind a subset of AP1/NF-E2 like sequences (Mignotte et al., 1989; Andrews et al., 1993a). NF-E2 has been shown to be an obligate heterodimer between the haematopoietic specific p45 protein and the ubiquitously expressed p18-maf protein, both members of the basic leucine zipper family of transcription factors (Igarashi et al., 1994; Andrews et al., 1993a and 1993b). NF-E2 binding sites have been identified in each of the LCR HS sites and shown to be particularly important in the enhancer function of the β-globin LCR HS2. The p18 or MafK subunit belongs to the small Maf family of bZIP transcription factors which have been shown to be able to switch from transcriptional repressors to activators depending on the dimer composition (Igarashi et al., 1994 and 1995; Kataoka et al., 1995). The p45-NF-E2 subunit belongs to the Cap'n'collar (CNC)-type bZIP protein family which includes Nrf1/LCR-F1 and Nrf2. P45-NF-E2 is abundantly expressed in haematopoietic cells and thought to play an important role in globin

gene expression and erythroid development. Homozygous disruption of the p45-NF-E2 gene in mice reveals a defective megakaryopoiesis and thrombopoiesis. However erythroid development was practically normal with only very mild red cell abnormalities and a slight decrease in erythroid haemoglobinisation (Shivdansani et al., 1995a and 1995b). It is possible that its genetic loss can be compensated by the erythroid co-expressed Nrf-1, Nrf-2 or the more recently identified novel member of the small bZip transcription factors Bach1 (Oyake et al., 1996). Deletion of both Nrf1/LCR-F1 genes demonstrated that this bZIP protein is essential for mesoderm formation in mice (Farmer et al., 1997). In addition, Nrf1/LCR-F1-wild type chimaeric mice were normal suggesting a non-cell autonomous role and implying that its function could be compensated by one of its family members or that it only regulates genes encoding intercellular molecules. Bach1 is expressed ubiquitously and has been shown to heterodimerise with the small MafK NF-E2 subunit and to bind and function as a transcriptional activator in transfection assays using reporter constructs containing the NF-E2 binding sites from the chicken β-globin 3' enhancer in chicken fibroblast cells (Oyake et al., 1996). This new member of the small Maf family of bZIP proteins might be a good candidate to play a role in the transcriptional regulation of the globin genes.

## 1.5 Developmental regulation of the human β-globin locus

Many of the *cis*- and *trans*-acting components involved in the erythroid specific transcriptional activation of the human β-globin genes have been identified and described in the previous section. However what role these components play and how the individual globin genes are regulated in their stage specific expression has been addressed by transgenic studies which will be discussed here.

*Autonomous versus competitive regulation*

When the expression patterns of the individual genes were tested in the absence of the LCR, the ε-globin gene was not expressed (Shih et al., 1990). The γ- and β-genes were expressed in embryonic and foetal/adult stages, albeit at very low levels compared to their murine counterparts (Magram et al., 1985; Townes et al., 1985; Kollias et al., 1986; Chada et al., 1986). Subsequently, expression studies of the individual globin genes linked in *cis* to the LCR showed partial reconstruction of correct developmental control. In transgenic mice and in differentiating embryonic bodies, a single embryonic ε-globin gene linked to the LCR is expressed, like its murine structural homologue εy, at a high level in primitive erythrocytes (Fraser, unpublished; Lindenbaum et al., 1990) even when linked to a partial LCR (Raich et al., 1990; Shih et al., 1990 and 1993). The γ-globin gene, like its murine structural homologue βH1, is expressed in the embryonic yolk sac. However, unlike the βH1-gene, the γ-gene is expressed during early foetal liver stages and is silenced around day 16 of development to remain silent in the adult stage (Dillon and Grosveld, 1991). This implies that both genes are regulated autonomously during development in which the proximal control elements of the ε- and γ-globin genes in combination with the LCR seem to be sufficient to direct correct developmental stage specific expression, including their silencing at the appropriate stages.

Several candidate binding sites for the action of repressor proteins have been identified, though the exact silencing mechanisms still remain unsolved. A putative silencer element is located in the upstream region of the embryonic ε-globin gene promoter between –177 and –392 of the transcriptional start site (Cao et al., 1989). Deletion of the silencer region results in the continued but very low expression of the ε-globin gene during fetal and adult stages

(Raich et al., 1992). Genetic data revealed that several point mutations that are spread throughout the promoter regions or a small deletion can alleviate the autonomous silencing of the foetal γ-globin genes. Several mutations map to the distal CCAAT box but have also been mapped to regions further upstream (reviewed in: Wood, 1993; Rochette, 1994; Pissard et al., 1996).

The human β-globin gene when directly linked to the LCR is prematurely and aberrantly expressed at high levels in embryonic red cells (Enver et al., 1990; Behringer et al., 1990). However, similar to its murine counterparts, the human gene is expressed at a high level in the definitive erythroid cells (Grosveld et al., 1987; Blom van Assendelft et al., 1989). Additional experiments in transgenic mice showed that the premature embryonic expression of the adult β-globin gene is abrogated by competition for the LCR by inserting either a γ- or an α-globin gene in between the LCR and the β-gene (Behringer et al., 1990; Enver et al., 1990; Hanscombe et al., 1991). This suggests that the genes compete for the activation function of the LCR. In experiments described by Hanscombe et al., 1991 the order of the genes in relation to the LCR was reversed, which altered the developmental patterns of gene expression. This indicates that the relative distance or gene order in relation to the LCR is an important parameter in determining the level of gene expression. Similar reciprocal changes in the level of gene expression were detected in non-deletion HPFH patients in which an increased level of γ-gene expression caused by a mutation in the γ-promoter sequence are accompanied by a reciprocal and equivalent decrease in the level of the distal β-globin gene in *cis* (Berry et al., 1992). These data strongly suggest that the genes compete with one another for the activation function of the LCR. During developmental periods of gene switching transgenic studies have shown that all the genes compete for LCR activity in a reciprocal fashion. The ordered arrangement of the genes provides the locus with a structural polarity in which a gene more proximal to the LCR has a competitive advantage to be activated by the LCR over a more distally located gene (Hanscombe et al., 1991). In addition to LCR proximity, also stage specific transcription factor environment has an important influence on determining the level of expression of the different globin genes during development. First of all the stage specific silencing of the ε- and γ-globin genes is thought to be due to stage specific repressor proteins. Also the experiments described by Hanscombe et al., 1991 clearly indicated a preference for γ- and β-gene expression in primitive and definitive erythroid cells, respectively. In the embryonic lineage the γ-gene is expressed at fairly high levels when placed at a distal position from the LCR in relation to the *cis*-linked β-gene however in reversed order the β-gene is completely silenced.

Parallels in the regulation of human globin gene switching has often been drawn with the chicken locus where competition of globin genes for a shared regulatory element was first observed between the chicken $\beta^A$- and ε-globin genes in transformed primitive and definitive erythroid cell lines (Choi and Engel, 1988). The shared enhancer was shown to function as an LCR-like element (Reitman et al., 1990) and is located between the adult $\beta^A$- and embryonic ε-globin genes at almost equal distance from both promoters (Choi and Engel, 1988). When linked as a single gene to the enhancer, the adult $\beta^A$-gene seems to be autonomously regulated at its correct developmental stage whereas the ε-gene is expressed in both transfected primitive as well as definitive cells (Choi and Engel, 1988). For correct embryonic-restricted expression of the ε-globin gene the presence of a *cis*-linked adult β-globin gene is required. Moreover mutation of individual promoter elements in the chicken β-globin gene leads to reciprocal changes in the expression of the ε- and β-globin genes (Foley and Engel, 1992). Duplication of the intergenic enhancer region by creating a tandem repeat at its original site resulted in the expression of both genes in definitive erythroid cells suggesting that the duplicated enhancers can function as separate entities activating both genes thereby removing

the competition (Choi and Engel, 1988; Foley and Engel, 1992).

Additional evidence for directional competition in globin gene expression comes from studies on the mouse β-globin locus (Skow et al., 1983). In adult mice, the *βmaj* and *βmin* globin genes are the only active adult β-globin genes, which are expressed at 80% and 20% of the total adult β-globin gene expression, respectively. Mice homozygous for an insertional mutation, which disrupts the adult *βmaj* gene, are anaemic and die perinatally (Shehee et al., 1993). However, mice homozygous for a 3.3 kb deletion of the proximal adult *βmaj* gene including its regulatory sequences revealed an that βmin globin synthesis in relation to α-globin gene expression was upregulated to about 50% in heterozygous and 75% in homozygous knockout mice, which survive normally (Skow et al., 1983).

Taken together, these observations lead to the proposal that the most important regulatory control in the developmental stage specific expression of the human globin genes is the suppression at the correct developmental stage of the embryonic ε- and foetal γ-globin genes (Dillon and Grosveld, 1991). However, the role of stage specific repressors does not account for the silencing of the γ-genes during early embryonic and the β-gene during early embryonic and fetal stages of human erythroid development, but this silencing rather seems to be the result of a normal location in the locus and competition with more proximal genes for the activation function of the LCR. This model suggests that activation of gene transcription is directed through direct LCR-gene interactions at all stages of development (Hanscombe et al., 1990; Figure 5). During embryonic stages the LCR preferably interacts and activates the most proximal embryonic ε-globin gene due to its competitive advantage. A switch from ε- to γ-gene expression is realised through negation of LCR ε-globin gene interactions allowing the LCR to interact and activate the γ-globin genes. The second switch in globin gene expression is realised through suppression of LCR γ-gene interactions in the adult, allowing for the β-gene to interact with the LCR (Figure 5). Polarity of LCR-gene interactions are due to relative proximity and thus the volumes in which the LCR-gene interactions operate, which determines the frequency of interactions while the transacting factor environment determines the strength of the interactions. Switching in gene expression during development in such a model would be smooth, keeping the total output of the β-locus constant and in balance with the output of the α-locus.

*Generation of mice carrying the full and mutant human β-globin locus*

*In vitro* ligation of two overlapping cosmids from the human β-globin locus produced a 70kb construct containing the full locus, which was introduced into transgenic mice (Strouboulis et al., 1992a and 1992b). The expression patterns of the human globin genes are quite similar to the patterns of expression discussed above and are illustrated for both the human and murine globin genes in Figure 6. The behaviour of the human β-locus in transgenic mice is in two aspects different from its behaviour in human. First, the more distal γ-globin genes are expressed at the earliest embryonic stages in the yolk sac at higher levels than the ε-globin gene. This can not be explained by simple competition, but may be caused by changes in the expression of at least one transcription factor which must have taken place in the process of fetal recruitment of the γ-genes in humans since mice do not have a typical fetal expressed globin gene (Strouboulis et al., 1992). A second difference is seen in the timing of γ- to β-gene switching normally occurring around birth in human, which in mice, however, takes place in the early foetal liver. Expression of the transgenic human globin genes is similar to the expression of the endogenous genes in humans, and characterised by two major gene switch events resulting in two periods of overlapping gene activity. Initially during embryonic stages in the primitive erythroid cells the ε- and γ-genes are co-expressed, whereas during early foetal stages the γ-and β-genes are simultaneously active. Detection

analysis of protein products of the different globin genes in single red blood cells during the periods of gene switching showed co-expression of ε and γ polypeptides in single primitive red cells and γ and β polypeptides in single definitive red cells, during early fetal stages of erythropoiesis (Strouboulis, 1994 PhD thesis; Fraser et al., 1993). This suggests that during switching a single red cell is not necessarily predetermined to express either gene singularly, but can express both genes during its life span. This is in agreement with the detected co-expression of fetal and adult haemoglobins in human fetal cells and adult F-cells (Stamatoyannopoulos and Nienhuis, 1987).

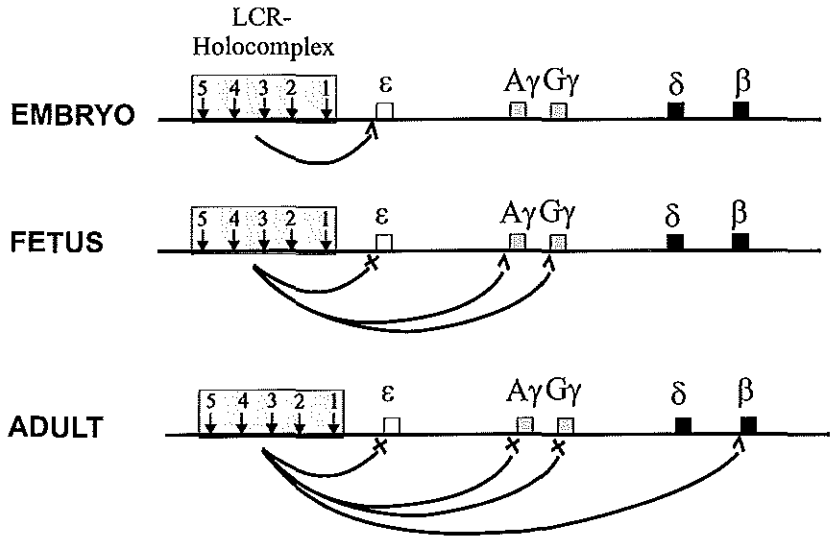



**Figure 5.** Model for regulation of gene expression in the human β-globin locus in the embryonic, fetal and adult stages of development (Grosveld et al., 1993).
An X indicates that the interaction between the LCR and the genes is blocked, which is thought to be caused by developmental stage-specific repressors binding to "most probably" promoter sequences. An arrow indicates a productive interaction. This Figure only indicates the LCR gene interactions which are preferred at each developmental stage. The γ-globin genes are already expressed during embryonic stages and the β-gene during fetal stages though at relatively low levels.

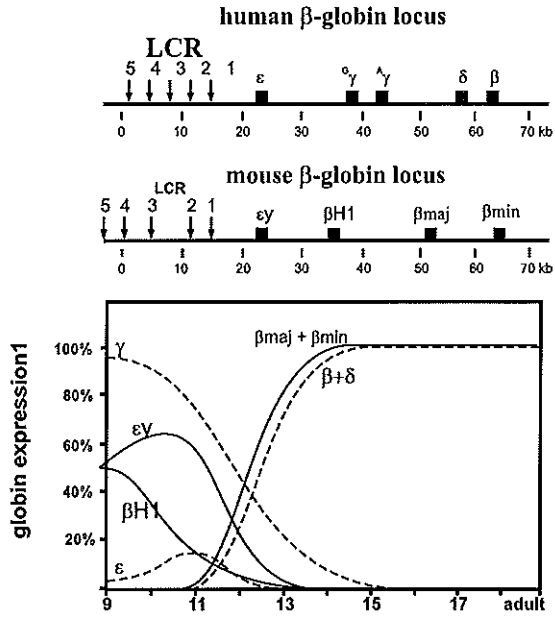



**Figure 6.** The organisation and developmental expression pattern of the human and mouse β-globin genes. The graph shows the developmental pattern of mRNA expression of the human β-globin genes in single copy transgenic mice in comparison to the mouse endogenous β-globin genes (Strouboulis et al., 1992).

When integrated into the murine genome, the human β-globin locus behaves quite similar to the endogenous mouse locus and to the human locus in its native environment in human red cells. Moreover, expression analysis, during mouse development, of two yeast artificial chromosomes (YACs) with a length of 150kb and 248kb and containing the full intact human β-globin locus, showed a very similar pattern as discussed above for the transgenic mice containing the 70kb human β-locus (Gaensler et al., 1993; Peterson et al., 1993a and 1993b). Therefore these single copy human globin transgenic mice are thought to represent good models for studying regulatory aspects of developmental human globin gene expression, and have been used as a model in three out of the four studies discussed in the following chapters of this thesis.

Similar methods were used to produce transgenic animals in which a genetically marked β-gene was located at two different relative positions in the otherwise intact full length human β-globin locus (Dillon et al., 1997). The marked β-gene (βm) in one construct was exchanged with the ε-gene (ML1) and in the other construct inserted upstream of the δ-gene (ML2). In this way the effect of gene position in the locus was directly tested and some additional conclusions were made concerning the activation of the genes by the LCR. Comparison of the

expression patterns during development indicated that all genes compete for the activation function of the LCR at all stages with an LCR-proximal gene having an advantage over a more distally located gene. This confirms the existence of polar competition in the context of the complete β-globin locus. When the βm gene is in the ε-position it competes more effectively with the fetal γ-genes than the ε-gene, suggesting no additional negative elements within the 2kb of flanking regions on either side of the β-gene which can silence the gene in primitive erythroid cells. This confirms that the β-gene is kept silent during embryonic and fetal stages of development by competition as a result of its distal position in the locus. Comparing the levels of expression produced by the marked (βm) and wild type β-genes (β) in adult blood showed that their relative expression levels changed with relative distances of the genes to the LCR. This result implies a more stochastic activation mechanism, which can be envisioned when direct interactions between the LCR and the genes are required for activation of transcription. The total output of both mutated loci, which is the sum of βm plus β levels, was similar to that of the wild type locus. This result indicates that the competition between the genes for activation by the LCR is balanced, and provides support for the fact that the LCR acts as a single larger entity or 'holocomplex' interacting with a single gene at a time rather than with multiple genes simultaneously.

### 1.6 Aim of the project

The β-globin locus adopts a fully active euchromatic configuration exclusively in the erythroid lineage. Once the chromatin structure is open, β-globin gene transcription is dependent on the interaction of transcription factors to at least three types of *cis*-regulatory elements located within the β-globin domain: the promoters including upstream promoter elements, the globin gene enhancers and the LCR. I have discussed a model in which the individual LCR HS elements act together as a single functional unit or 'holocomplex', which subsequently interacts preferentially with the most proximal transcriptionally competent gene via looping. However two additional alternative mechanisms have been considered for the transcriptional activation of the human β-globin genes. The three models are described as the accessibility model, the scanning or tracking model, and the looping model and have been considered for the action of classical enhancer elements on nearby promoters. The models are schematically illustrated in Figure 7. The accessibility model implies that the only function of the LCR is to set up and maintain the entire locus in an open chromatin structure. The genes in the locus would then simply be accessible to the binding of transcription factors and thus be transcribed in a stochastic fashion (Groudine and Weintraub, 1982; Martin et al., 1996). Since the dimensional aspects of the locus would not play a role in the regulation of the genes, an additional mechanism called transcriptional interference was proposed, which states that transcriptional activity of an upstream gene in the locus can interfere with the transcriptional activity of a downstream gene. Globin gene switching would be a result of the gradual change in the frequency of the on and off response of the transcriptional activity of a gene as a result of the changing transacting factor environment. The scanning or tracking model proposes that the LCR functions as an entry site for the binding of the transcriptional machinery, which would scan along the DNA. Activation is unidirectional in that it would preferably activate the first promoter that it scans into which upon the right transacting factor environment is rendered transcriptionally active (Herendeen et al., 1992; Tuan et al., 1992). In the looping model, transcriptional activation requires that the LCR participates in direct chromatin interactions with gene proximal regulatory sequences of a single gene (Ptashne, 1988; Bickel and Pirotta, 1990; Mueller storm et al., 1989; Folley and Engel, 1992; Dillon et al., 1995).

Since the looping model predicts that the LCR gene interactions operate in a three-dimensional manner, transcriptional activation will heavily depend on the relative distance to allow the LCR to interact stochastically with a gene that is rendered transcriptionally active by the proper transacting factor environment.

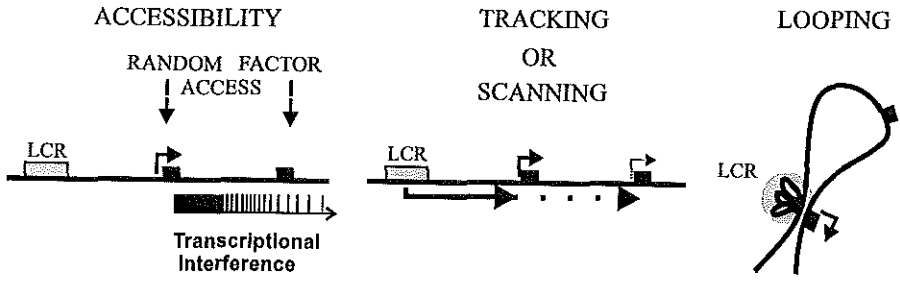



**Figure 7.** Three models that have been proposed to provide an explanation for how the distant β-globin LCR activates the individual β-globin genes during development.
How the models mechanistically explain the activation effect of the β-LCR over long distances is discussed in the text above. The filled squares indicate genes and the arrow indicates a transcriptionally active gene. The gradual increase in space between the small vertical lines underneath the accessibility model indicates that the proposed transcriptional interference caused by transcriptional activity of an upstream gene decreases with distance.

The occurrence of competition in the regulation of the β-globin cluster of genes implies that at least part of the developmental specificity provided by the LCR is the result of differential interaction with the genes. Promoter competition for a shared enhancer region has been observed in the regulation of numerous other genes and seem to represent more of a general principle in the action of an enhancer when shared by multiple genes. Enhancer sharing, selectivity and competition have been reported as general mechanisms occurring between the *Hoxb4* and *Hoxb5* genes and the *Hoxb5* and *Hoxb6* genes, in order to generate the proper spatial and temporal patterns of *Hox* gene expression (Sharpe et al., 1998). Temporal co-linearity was challenged when *Hoxd9* or *Hoxd11* were transposed to a similar 5' position in the *Hoxd* cluster, indicating the existence of enhancer sharing between neighbouring genes (van der Hoeven et al., 1996). These mechanisms have been proposed to underlie the maintenance of the clustered organisation of the *Hox* genes throughout evolution. Furthermore, gene competition for a common set of enhancers has been proposed to explain the reciprocal imprinting of the murine *H19* and *Igf2* genes (Bartolomei and Tilghman, 1992). This has been elegantly confirmed through the addition of an extra set of enhancers in the

intergenic region through homologous recombination which relieved the silencing of the maternally inherited *Igf2* allele implying a relief of competition (Webber et al., 1998). Thus it seems that a similar competition mechanism as we have observed in the human β-globin locus underlies the regulation of the *Hox*, *H19* and *Igf2* genes and probably many other gene clusters as well. Trying to elucidate what molecular mechanisms underlie gene competition for common regulatory elements is thus a matter of general interest.

During my PhD studies, I was mainly concerned with the experimental design of how we would be able to distinguish between the three transcriptional activation models. First we addressed how transcriptional activation dynamics regulate globin gene transcription during stages of globin gene switching and what the implications could be for how the LCR activates the individual globin genes. A key question is whether only a single gene is active at a time or multiple genes simultaneously. Two different labs have proposed that the LCR can activate multiple genes simultaneously, based on the presence of multiple mRNA species within single erythroid cells (Furakawa et al., 1994 and 1995) and on the presence of accessibility to restriction endonucleases that cleave both γ-gene promoters at a high percentage in erythroid nuclei (Bresnick et al., 1994; Furakawa et al., 1995). Detection of globin mRNA or polypeptides are not reliable indicators of active gene transcription, due to their long half lives. Neither is restriction nuclease accessibility since its is not necessarily linked to transcriptional activity. We considered that the only way to distinguish between single versus multigene activity, is by looking within single cells at a relatively short lived event which is closely linked to the transcriptional activity of a gene. Taking these considerations into account we developed a sensitive fluorescent *in situ* hybridisation (FISH) technique which allows for detection of gene specific intron RNA sequences (RNA FISH) in the nucleus of a single red cell at the site of transcription (Chapter 2). Intron sequences can be found in nascent transcripts and in full length primary transcripts, which are rapidly spliced out to form short lived intron lariats. These intron containing RNA intermediates are both temporally and spatially associated with the transcriptional event, and their short half lives make them good indicators of concurrent or very recent transcriptional activity. Throughout our studies we made use of this RNA FISH technique in order to study the transcription dynamics of globin gene competition in transgenic mice. The results in Chapter 2 show that gene transcription in the human β-globin locus is an all or nothing event in that a gene is either fully active or silent. Only a single gene is active at a time in *cis* and multigene expression in single cells is a result of alternate transcription of single genes. As described in Chapter 5 we extend these analyses by following a novel kinetic approach in which we use transcriptional inhibition and release in conjunction with RNA FISH. This study confirmed that multiple globin-gene primary transcript signals in *cis* represent a transition between alternating periods of single gene transcription rather than co-initiation of multiple genes in the locus. In Chapter 3 we studied the role of EKLF in gene competition between the γ and β genes for LCR activation in compound EKLF knockout and human β-globin locus transgenic mice, to further unravel the mechanism of developmental regulation. In Chapter 4 we augmented our understandings of the role of EKLF during development, and it is shown that EKLF rather than SP1 is the activator that binds the CACC element in LCR HS site 3. I will discuss how the results presented in this thesis in combination with additional recent advances and previously obtained results argue in favour of the LCR holocomplex or looping model.

## REFERENCES

Almer, A. and Hörz, W. (1986). Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. *EMBO J.*, **5**, 2681-2687.

Almer, A., Rudolgh, A. Hinnen, A. and Hörz, W. (1986). Removal of positioned nucleosomes from yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. *EMBO J.*, **5**, 2689-2696.

Andrews, N.C. and Faller, D.V. (1991). A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucl. Acids Res.*, **19**, 2499.

Andrews, N.C., Kotkow, K., Ney, P., Erdjument-Bromage, H., Tempst, P. and Orkin, S.H. (1993a).The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the *V-maf* oncogene. *Proc. Natl. Acad. Sci. USA*, **90**, 11488-11492.

Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. and Orkin, S.H. (1993b). Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature*, **362**, 722-728.

Antoniou, M., deBoer, E., Habets, G. and Grosveld, F. (1988). The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *EMBO J.*, **7**, 377-384.

Antoniou, M. and Grosveld, F. (1990). The beta-globin gene dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev.*, **4**, 1007-1013.

Antoniou, M. (1991). Induction of erythroid-specific expression in murine erythroleukemia (MEL) cell lines. *Med. Mol. Biol.*, **7**, 421-434.

Antoniou, M., deBoer, E., Spanopoulou, E., Imam, A. and Grosveld, F. (1995). TBP binding and the rate of transcription initiation from the human β-globin gene. *Nucl. Acids Res.*, **23**, 3473-3480.

Armstrong, J.A. and Emerson, B.M. (1996). NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 *in vitro*. *Mol. Cell. Biol.*, **16**, 5634-5644.

Banerji, J., Rusconi, S. and Schaffner, W. (1981). Expression of beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 299-308.

Baniahmad, A., Kohne, A.C. and Renkawitz, R. (1992). A transferable silencing domain is present in the thyroid hormone receptor, in the v-erbA onco-gene product and in the retinoic acid receptor. *EMBO J.*, **11**, 1015-1023.

Bannister, A.J. and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. *Nature*, **384**, 641-643.

Baron, M.H. and Maniatis, T. (1986). Rapid reprogramming of globin gene expression in transient heterokaryons. *Cell*, **46**, 591-602.

Baron, M.H. (1997). Transcriptional control of globin gene switching during vertebrate development. *Biochem. Bioph. Acta*, **1351**, 51-72.

Bartolomei, M.S. and Tilghman, S.M. (1992). Parental imprinting of mouse chromosome 7. *Semin. Dev. Biol.*, **3**, 107-117.

Behringer, R.R., Hammer, R.E., Brinster, R.L., Palmiter, R.D. and Townes, T.M. (1987). Two 3' sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. *Proc. Natl. Acad. Sci. USA*, **84**, 7056-7060.

Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L. and Townes, T.M. (1990). Human gamma- to beta-globin gene switching in transgenic mice. *Genes Dev.*, **4**, 380-389.

Benoist, C. and Chambon, P. (1981). *In vivo* sequence requirements of the SV40 early promoter region. *Nature*, **290**, 304-310.

Berry, M., Dillon, N. and Grosveld, F. (1992). A single point mutation is the cause of the Greek form of hereditary persistence of foetal haemoglobin. *Nature*, **358**, 499-502.

Bickel, S. and Pirrotta, V. (1990). Self-association of the *Drosophila* zeste protein is responsible for transvection effects. *EMBO J.*, **9**, 2959-2967.

Bieker, J.J. (1996). Isolation, genomic structure, and expression of human erythroid Krüppel-like factor (EKLF). *DNA & Cell Biol.*, **15**, 347-352.

Blake, C.C.F. (1978). Do genes-in-pieces imply proteins-in-pieces? *Nature*, **273**, 267.

Blom van Assendelft, G., Hanscombe, O., Grosveld, F. and Greaves, D.R. (1989). The β-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell*, **56**, 969-977.

Bodine, D.M. and Ley, T.J. (1987). An enhancer element lies 3' to the human $^{A}$gamma globin gene. *EMBO J.*, **6**, 2997-3004.

Bonifer, C., Vidal, M., Grosveld, F. and Sippel, A.E. (1990). Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *EMBO J.*, **9**, 2843-2848.

Bonifer, C., Yannoutsos, N., Krüger, G., Grosveld, F. and Sippel, A.E. (1994). *Nucl. Acids Res.*, **22**, 4202-4210.

Boyes, J. and Felsenfeld, G. (1996). Tissue-specific factors additively increase the probability of the all-or-none formation of a hypersensitive site. *EMBO J.*, **15**, 2496-2507.

Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R. and Nasmyth, K. (1985). Characterisation of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell*, **41**, 41-48.

Bresnick, E.H., Bustin, M. Marsaud, V., Richard-Foy, H. and Hager, G.L. (1992). The transcriptionally active

MMTV promoter is depleted of histone H1. *Nucl. Acids Res.*, **20**, 273-278.
Bresnick, E.H. and Felsenfeld, G. (1994). Dual promoter activation by the human β-globin locus control region. *Proc. Natl. Acad. Sci. USA*, **91**, 1314-1317.
Brethnach, R. and Chambon, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.*, **50**, 349-383.
Briegel, K., Lim, K.C., Plank, C., Beug, H., Engel, J.D. and Zenke, M. (1993). Ectopic expression of a conditional GATA-2/estrogen receptor chimera arrests erythroid differentiation in a hormone-dependent manner. *Genes Dev.*, **7**, 1097-1109.
Bungert, J. and Engel, J.D. (1996). The role of transcription factors in erythroid development. *Ann. Med.*, **28** (1), 47-55.
Bunn, H.F. and Forget, B.G. (1986). Haemoglobin: Molecular, Genetic and Clinical Aspects. Philadelphia, PA: WB Saunders.
Buratowski, S. (1994). The basics of basel transcription by RNA polymerase II. *Cell*, **77**, 1-3
Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B. and Kornberg, R.D. (1996). RSC, an essential, abundant chromatin-remodelling complex. *Cell*, **87**, 1249-1260.
Cao, S., Gutman, P.D., Dave, H.P.G. and Schechter, A.J. (1989). Identification of a transcriptional silencer in the 5'-flanking region of the human ε-globin gene. *Proc. Natl. Acad. Sci. USA*, **86**, 5306-5309.
Cao, A., Galanello, R., Rosatelli, M.C. (1994) Genotype-phenotype correlations in ß-thalassaemias. *Blood Reviews*, **8**, 1-12.
Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R. and Townes, T.M. (1994a). Multiple elements in human β-globin locus control region 5' HS2 are involved in enhancer activity and position-independent, transgene expression. *Nucl. Acids Res.*, **22**, 1006-1011.
Caterina, J.J., Donze, D., Sun, C.W., Ciavatta, D.J. and Townes, T.M. (1994b). Cloning and functional characterisation of LCR-F1: a bZip transcription factor that activates erythroid-specific, human globin gene expression. *Nucl. Acids Res.*, **12**, 2383-2391.
Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E. and Constantini, F. (1985). Specific expression of a foreign beta-globin gene in erythroid cells of transgenic mice. *Nature*, **314**, 377-380.
Chada, K., Magram, J. and Constantini, F. (1986). An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature*, **319**, 685-689.
Chakravarti, D., LaMorte, V.J., Nelson, M.C., Nakajima, T., Schulman, I.G., Juguilon, H. Montminy, M. and Evans, R.M. (1996). Role of CBP/P300 in nuclear receptor signalling. *Nature*, **383**, 99-103.
Chan, J.Y., Han, X.L. and 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.
Chao, M.V., Mellon, P.., Charnay, P., Maniatis, T. and Axel, R. (1983). The regulated expression of ß-globin genes introduced into mouse erythroleukemia cells. *Cell*, **32**, 483-493.
Chavez, S., Candau, R., Truss, M. and Beato, M. (1995). Constitutive repression and nuclear factor I-dependent hormone activation of the mouse mammary tumor virus promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **15**, 6987-6998.
Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988). Human CCAAT-binding proteins have heterologous sub-units. *Cell*, **55**, 11-24.
Choi, O.R.B. and Engel, J.D. (1988). Developmental regulation of beta-globin gene switching. *Cell*, **55**, 17-26.
Clarke, D.J., O'Neill, L.P. and Turner, B.M. (1993). Selective use of H4 acetylation sites in the yeast *Saccharomyces cerevisiae*. *Biochem. J.*, **294**, 557-561.
Cockerill, P.N. and Garrard, W.T. (1986). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs net to the enhancer in a region containing topoisomerase II sites. *Cell*, **44**, 273-282.
Collins, F.S. and Weissman, S.M. (1984). The molecular genetics of human haemoglobin. *Prog. Acid Res. Mol. Biol.*, **31**, 315-462.
Collis, P., Antoniou, M. and Grosveld, F. (1990). Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J.*, **9**, 233-240.
Constantini, F. and Lacy, E. (1981). Introduction of a rabbit β-globin gene into the mouse germ line. *Nature*, **294**, 92-94.
Côté, J., Quinn, J., Workman, J.L. and Peterson, C.L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, **265**, 53-60.
Craddock, C.F., Vyas, P.V., Sharpe, J.A., Ayyub, H., Wood, W.G., and Higgs, D.R. (1995). Contrasting effects of α and β globin regulatory elements on chromatin structure may be related to their different chromosomal environments. *EMBO J*, **14**, 1718-1726.
Craig, J.E., Rochette, J., Fisher, C.A., Weatherall, D.J., Marc, S., Lathrop, G.M., Demenais, F. and Thein, S. (1996). Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nat. Genet.*, **12**, 58-64.

Crossley, M., Whitelaw, E., Perkins, A., Williams, G., Fujiwara, Y. and Orkin, S.H. (1996). Isolation and characterisation of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells. *Mol. Cell. Biol.*, **16**, 1695-1705.

Dale, T.C., Krnacik, M.J., Schmidhauser, C., Yang, C.L., Bissel, M.J. and Rosen, J.M. (1992). High-level expression of the rat whey acidic protein gene is mediated by elements in the promoter and 3' untranslated region. *Mol. Cell. Biol.*, **12**, 905-914.

deBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988). The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. *EMBO J.*, **7**, 4203-4212.

Diaz, P., Cado, D. and Winato, A. (1994). A locus control region in the T cell receptor alpha/delta locus. *Immunity*, **1**, 207-217.

Dickerson, R.E. and Geiss, I. (1983). Hemoglobin. London: The Benjamin Cummings Publishing company, inc.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell*, **32**, 695-706.

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

Dillon, N. and Grosveld, F. (1993). Transcriptional regulation of multigene loci: multilevel control. *TIG*, **9**, 134-137.

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell.*, **1**, 131-139.

Donze, D., Townes, T.M. and Bieker, J.J. (1995). Role of erythroid Krüppel-like factor in human γ- to β-globin gene switching. *J. Biol. Chem.*, **270**, 1955-1959.

Donze, D., Jeancake, P.H. and Townes, T.M. (1996). Activation of δ-globin gene expression by erythroid krüpple-like factor: A potential approach for gene therapy of sickle cell disease. *Blood*, **88**, 4051-4057.

Driel, R. and Otte, A.P. (Edited by) (1997). Nuclear organisation, chromatin structure and gene expression. *Oxford university press.*

Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989). Gamma delta beta-thalassaemia due to a *de novo* mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proc. Natl. Acad. Sci. USA*, **86**, 7470-7474.

Dynan, W. (1989). Modularity in promoters and enhancers. *Cell*, **58**, 1-4.

Earnshaw, W.C. and Heck, M.M.S. (1985). Localisation of topoisomerase II in mitotic chromosomes. *J. Cell Biol.*, **100**, 1716-1725.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissmann, S.M. and et al. (1980). The structure and evolution of the human β-globin gene family. *Cell*, **21**, 653.

Elgin, S.C.R. (1988). The formation and function of DNase I hypersensitive sires in the process of gene activation. *J. Biol. Chem.*, **263**, 19259-19262.

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

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

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

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

Ess, K.C., Whitaker, T.L., Cost, G.J., Witte, D.P., Hutton, J.J. and Aronow, B.J. (1995). A central role for a single c-Myb binding site in a thymic locus control region. *Mol. Cell. Biol.*, **15**, 5707-5715.

Evans, T. and Felsenfeld, G. (1989). The erythroid-specific transcription factor Eryf1: A new finger protein. *Cell*, **58**, 877-885.

Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. *Nature*, **355**, 219-224.

Felsenfeld, G. (1996). Chromatin unfolds. *Cell*, **86**, 13-19.

Felsenfeld, G., Boyes, J., Chung, J., Clark, D. and Studitsky, V. (1996). Chromatin structure and gene expression. *Proc. Natl. Acad. Sci. USA*, **93**, 9384-9388.

Feng, W.C., Southwood, C.M. and Bieker, J.J. (1994). Analyses of β-thalassaemia mutant DNA interactions with erythroid Krüppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J. Biol. Chem.*, **269**, 1493-1500.

Foley, K.P. and Engel, J.D. (1992). Individual stage selector element mutations lead to reciprocal changes in β-

vs. ε-globin gene transcription: genetic confirmation of promoter competition during globin gene switching. *Genes Dev.*, **6**, 730-744.

Fong, T.C. and Emerson, B.M. (1992). The erythroid-specific protein cGATA-1 mediates distal enhancer activity through a specialised β-globin TATA box. *Genes Dev.*, **6**, 521-532.

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

Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. (1987). Evidence for a locus activator region. *Nucl. Acids Res.*, **15**, 10159-10177.

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

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

Fraser, P.J. and Curtis, P.J. (1987). Specific pattern of gene expression during induction of mouse erythroleukemia cells. *Genes Dev.*, **1**, 855-861.

Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990). DNaseI hypersensitive sites 1,2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucl. Acids Res.*, **18**, 3503-3508.

Fraser, P., Pruzina, S., Antoniou, M. and 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.

Furukawa, T., Zitnik, G., Leppig, K., Papayannopoulou, T. and Stamatonyannopoulos, G. (1994). Coexpression of γ and β globin mRNA in cells containing a single human β globin locus: Results from studies using single-cell reverse transcription polymerase chain reaction. *Blood*, **83**, 1412-1419.

Furukawa, T., Navas, P.A., Josephson, B.M., Peterson, K.R., Papayannopoulou, T. and Stamatonyannopoulos, G. (1995). Coexpression of ε, $^{G}\gamma$ and $^{A}\gamma$ globin mRNA in embryonic red blood cells from a single copy β-YAC transgenic Mouse. *Blood Cells Mol. Dis.*, **21**, 168-178.

Gaensler, K.M., Kitamura, M. and Kan, Y.W. (1993). Germ-line transmission and developmental regulation of a 150-kb yeast artificial chromosome containing the human beta-globin locus in transgenic mice. *Proc. Natl. Acad. Sci. USA*, **90**, 11381-11385.

Garel, A. and Axel, R. (1976). Selective digestion of transcriptionally active ovalbumin genes from oviduct nuclei. *Proc. Natl. Acad. Sci. USA*, **73**, 3966-3970.

Gasser, S.M. and Laemmli, U.K. (1986a). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster. *Cell*, **46**, 521-530.

Gasser, S.M., Laroche, T., Falquet, J., Boy de la Tour, E. and Laemmli, U.K. (1986b). Metaphase chromosome structure. Involvement of topoisomerase II. *J. Mol. Biol.*, **188**, 613-629.

Gilbert, W. (1978). Why genes in pieces? *Nature*, **271**, 501.

Gill, G. (1994). Transcriptional initiation. Taking the initiative. *Curr. Biol.*, **4**, 374-376.

Gong, Q.H., Stern, J. and Dean, A. (1991). Transcriptional role of a conserved GATA-1 site in the human ε-globin gene promoter. *Mol. Cell. Biol.*, **11**, 2558-2566.

Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A. and Ruddle, F.H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. USA*, **77**, 7380-7384.

Graziano, V., Gerchman, S.E., Schneider, D.K. and Ramakrishnan, V. (1994). Histone H1 is located in the chromatin 30-nm filament. *Nature*, **368**, 351-354.

Greaves, D.R., Wilson, F.D., Lang G. and Kioussis, D. (1989). Human CD2 3'-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. *Cell*, **56**, 979-986.

Gribnau, J., de Boer, E., Trimborn, T., Wijgerde, M., Milot, E., Grosveld, F. and Fraser, P. (1998). Chromatin interaction mechanism of transcriptional control. *Submitted.*

Gross, D.S. and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.*, **57**, 159-197.

Grosschedl, R. and Birnstiel, M.L. (1980a). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. *Proc. Natl. Acad. Sci. USA*, **77**, 1432-1436.

Grosschedl, R. and Birnstiel, M.L. (1980b). Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene transcription *in vivo*. *Proc. Natl. Acad. Sci. USA*, **77**, 7102-7106.

Grosveld, G.C., deBoer, E., Shewmaker, C.K. and Flavell, R.A. (1982). DNA sequences necessary for transcription of the rabbit beta-globin gene *in vivo*. *Nature*, **295**, 120-126.

Grosveld, F., Blom van Assendelft, G., Greaves, D.R. and Kollias, G. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell*, **51**, 975-985.

Grosveld, F., Dillon, N., and Higgs, D. (1993). The regulation of human globin gene expression. *Chapter 2 in*

*"Baillière's Clinical Haematology"*, **6**, 31-55.
Groudine, M. and Weintraub, H. (1982). Propagation of globin DNaseI -hypersensitive sites in absence of factors required for induction: A possible mechanism for determination. *Cell*, **30**, 131-139.
Groudine, M., Kohwi-Shigematsu, T., Gelinas, R., Stamatoyannopoulos, G. and Papayannopoulou, T. (1983). Human fetal to adult haemoglobin switching: changes in chromatin structure of the β-globin gene locus. *Proc. Natl. Acad. Sci. USA*, **80**, 7551-7555.
Grunstein, M. (1990). Histone function in transcription. *Annu. Rev. Cell Biol.*, **6**, 643-678.
Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature*, **389**, 349- 352.
Gumucio, D.L., Rood, K.L., Blanchard-McQuate K.L. et al. (1991). Interaction of Sp1 with the human γ-globin promoter: binding and transactivation of normal and mutant promoters. *Blood*, **78**, 1853-1863.
Han, K. and Manley, J.L. (1993). Transcriptional repression by the *Drosophila* even-skipped protein: definition of a minimal repression domain. *Genes Dev.*, **7**, 491-503.
Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991). Importance of globin gene order for correct developmental expression. *Genes Dev.*, **5**, 1387-1394.
Hardison, R. (1991). Evolution of globin gene families. Chapter 13 in "Evolution at the molecular level" edited by Selander, R.K., Clark, A.G. and Whittam, T.S., 272-333.
Hardison, R. and Miller, W. (1993). Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol. Biol. Evol.*, **10**, 73-102.
Hardison, R.C. (1996). A brief history of haemoglobins: Plant, animal, protist and bacteria. *Proc. Natl. Acad. Sci. USA*, **93**, 5675-5679.
Hebbes, T.R., Clayton, A.L., Thorne, A.W. and Crane-Robinson, C. (1994). Core histone hyperacetylation co-maps with generalised DNase I sensitivity in the chicken β-globin chromosomal domain. *EMBO J.*, **13**, 1823-1830.
Herendeen, D.R., Kassavetis, G.A. and Geiduschek, E.P. (1992). A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science*, **256**, 1298-1303.
Higgs, D.R., Vickers, M.A., Wilkie, A.O., Pretorius, I.M., Jarman, A.P. and Weatherall, D.J. (1989). A review of the molecular genetics of the human α-globin gene cluster. *Blood*, **73**, 1081-1104.
Higgs, D.R. (1993). The haemoglobinopathies. *Chapter 5 in "Baillière's Clinical Haematology*, **6**, 117-150.
Ho, I.C. and Leiden, J.M. (1990a). Regulation of the human T-cell receptor alpha gene enhancer: multiple ubiquitous and T-cell-specific nuclear proteins interact with four hypomethylated enhancer elements. *Mol. Cell Biol.*, **10**, 4720-4727.
Hofer, E., Hofer-Warbinek, R. and Darnell J.E. (1982). Globin RNA transcription: A possible termination site and demonstration of transcriptional control correlated with altered chromatin structure. *Cell*, **29**, 887-893.
Huang, H.C. and Cole, R.D. (1984). The distribution of H1 histone is nonuniform in chromatin and correlates with different degrees of condensation. *J. Biol. Chem.*, **259**, 14237-14242.
Huang, J.D., Schwyter, D.H., Shirokawa, J.M. and Courey, A.J. (1993). The interplay between multiple enhancer and silencer elements defines the pattern of *decapentaplegic* expression. *Genes Dev.*, **7**, 694-704.
Hug, B.A., Moon, A.M. and Ley, T.J. (1992). Structure and function of the murine β-globin locus control region 5' HS-3. *Nucl. Acids Res.*, **20**, 5771-5778.
Hug, B.A., Wesselschmidt, R.L., Fiering, S., Bender, M.A., Epner, E., Groudine, M. and Ley, T.J. (1996). Analysis of mice containing a targeted deletion of β-globin locus control region 5' hypersensitive site 3. *Mol. Cell. Biol.*, **16**, 2906-2912.
Iarovaia, O., Hancock, R., Lagarkova, M., Miassod, R. and Razin, S.V. (1996). Mapping of genomic DNA loop organisation in a 500-kilobase region of the *drosophila* X chromosome by the topisomerase II-mediated DNA loop excision protocol. *Mol. Cell. Biol.*, **16**, 302-308.
Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M. and Yamamoto. M. (1994). Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature*, **367**, 568-572.
Igarashi, K., Itoh, K., Motohashi, H., Hasashi, N., Matazaki, Y., Nakauchi, H., Nishizawa, M. and Yamamoto, M. (1995). Activity and expression of murine small Maf family protein MafK. *J. Biol. Chem.*, **270**, 7615-7624.
Ikuta, T. and Kan, Y.W. (1991). *In vivo* protein-DNA interactions at the β-globin gene locus. *Proc. Natl. Acad. Sci. USA*, **88**, 10188-10192.
Imbalzano, A.N., Kwon, H., Green, M.R. and Kingston, R.E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature*, **370**, 481-485.
Jackson, J.R. and Benyajati, C. (1993). DNA-histone interactions are sufficient to position a single nucleosome juxtaposing *Drosophila* Adh adult enhancer and distal promoter [published erratum appears in Nucleic Acids Res. 1993 May 11; 21 (9):2286]. *Nucl. Acids Res.*, **21**, 957-967.
Jane, S.M., Ney, P.A., Vanin, E.F., Gumucio, D.L. and Nienhuis. A.W. (1992). Identification of stage selector element in the human γ-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the β-promoter. *EMBO J.*, **11**, 2961-2969.

Jane, S.M., Nienhuis, A.W. and Cunningham, J.M. (1995). Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein. *EMBO J.*, **14**, 97-105.

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

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S.T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.*, **14**, 116-127.

Jeffreys, A.J. and Flavell, R.A. (1977). The rabbit ß-globin gene contains a large insert in the coding sequences. *Cell*, **12**, 1097-1108.

Jeppesen, P. and Turner, B.M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell*, **74**, 281-289.

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

Jimenez, G., Griffiths, S.D., Ford, A.M., Greaves, M.F. and Enver, T. (1992). Activation of the β-globin locus control region precedes commitment to the erthroid lineage. *Proc. Natl. Acad. Sci. USA*, **89**, 10618-10622.

Johnson, P.F. and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.*, **58**, 799-839.

Jones, B.K., Monks, B.R., Liebhaber, S.A. and Cooke, N.E. (1995). The human growth hormone gene is regulated by a multicomponent locus control region. *Mol. Cell. Biol.*, **15**, 7010-7021.

Jonsson, Z.O. and Hubscher, U. (1997). Proliferating cell nuclear antigen: More than a clamp for DNA polymerases. *BioEssays*, **19**, 967-975.

Kamei, Y., Xu, L., Heinzel, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S.C., Heyman, R.A., Rose, D.W., Glass, C.K. and Rosenfeld, M.G. (1996). A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell*, **85**, 403-414.

Karpov, V.L., Preobrazhenskaya, O.V. and Mirzabekov, A.D. (1984). Chromatin structure of *hsp 70* genes, activated by heat shock: selective removal of histones from the coding region and their absence from the 5' region. *Cell*, **36**, 423-431.

Kataoka, K., Igarashi. K., Itoh, K., Fujiwara, K.T., Noda, M., Yamamoto, M. and Nishizawa, M. (1995). Small Maf proteins heterodimerize with Fos and potentially act as competitive repressors of NF-E2 transcription factor. *Mol. Cell. Biol.*, **15**, 2180-2190.

Kaul, D.K., Fabry, M.E., Nagel, R.L. (1996). The pathophysiology of vascular obstruction in the sickle syndromes. *Blood reviews*, **10**, 29-44.

Kelman, Z. (1997). PCNA: Structure function and interaction. *Oncogene*, **14**, 629-640.

Kim, L.J., Nikolov, D.B. and Burley, S.K. (1993). Co-crystal structure of TBP recognising the minor groove of a TATA element. *Nature*, **365**, 520-527.

Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P.B. (1993). Crystal structure of yeast TBP/TATA-box complex. *Nature*, **365**, 512-520.

Kioussis, D., Vanin, E., deLange, T., Flavell, R.A. and Grosveld, F.G. (1983). Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature*, **306**, 662-666.

Kioussis, D. and Festenstein, R. (1997). Locus control regions: Overcoming heterochromatin-induced gene inactivation in mammals. *Curr. Opin. Genet. Dev.*, **7**, 614-619.

Kollias, G., Wrighton, N., Hurst, J. and Grosveld, F. (1986). Regulated expression of human A gamma-, beta-, and hybrid gamma beta-globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell*, **46**, 89-94.

Kollias, G., Hurst, J., deBoer, E. and Grosveld, F. (1987). The human beta-globin gene contains a downstream developmental specific enhancer. *Nucl. Acids Res.*, **15**, 5739-5747.

Kong, S., Bohl, D., Li, C. and 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.

Kornberg, R.D. and Thomas, J.O. (1974). Chromatin structure: oligomers of the histones. *Science*, **184**, 865-868.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868-871.

Kornberg, R.D. (1977). Structure of chromatin. *Annu. Rev. Biochem.*, **46**, 931-954.

Kovacs, B.J. and Butterworth, P.H. (1986a). DNA sequences which influence the selection and efficient utilisation of transcriptional start sites of a eukaryotic gene by RNA polymerase II *in vitro* and *in vivo*. *Biosci Rep*, **6**, 937-944.

Kovacs, B.J. and Butterworth, P.H. (1986b). The effect changing the distance between the TATA-box and cap site by up to three base pairs on the selection of the transcriptional start site of a cloned eukaryotic gene *in vitro* and *in vivo*. *Nucl. Acids Res.*, **14**, 2429-2442.

Krude, T. and Elgin, S.C.R. (1996). Chromatin: Pushing nucleosomes around. *Curr. Biol.*, **6**, 511-515.

Kulozik, A.E., Bail, S., Bellan-Koch, A., Bartram, C.R., Kohne, E. and Kleihauer, E. (1991). The proximal

element of the beta-globin locus control region is not functionally required *in vivo*. *J. Clin. Invest.*, **87**, 2142-2146.

Kuo, M.H., Brownell, J.E., Sobel, R.E., Ranalli, T.A., Cook, R.G., Edmondson, D.G., Roth, S.Y. and Allis, C.D. (1996). Transcription-linked acetylation by Gcn5p of histone H3 and H4 at specific lysines. *Nature*, **383**, 269-272.

Kwon, H., Imbalzano, A., Khavari, P.A., Kingston, R.E. and Green, MR. (1994). *Nature*, **370**, 477-481.

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

Lee, D.Y., Hayes, J.J., Pruss, D. and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to DNA. *Cell*, **72**, 73-84.

Licht, J.D., Hanna-Rose, W., Reddy, J.C., English, M.A., Ro, M., Grossel, M., Shaknovich, R. and Hansen, U. (1994). Mapping and mutagenesis of the amino-terminal transcriptional repression domain of the *Drosophila* Krüppel protein. *Mol. Cell. Biol.*, **14**, 4057-4066.

Lindenbaum, M. and Grosveld, F. (1990). An *in vitro* globin gene switching model based on differentiated embryonic stem cells. *Genes Dev.*, **4**, 2075-2085.

Liu, Q., Bungert, J. and Engel, J.D. (1997). Mutation of gene-proximal regulatory elements disrupts human ε-, γ-, and β-globin expression in yeast artificial chromosome transgenic mice. *Proc. Natl. Acad. Sci. USA*, **94**, 169-174.

Loe, P.V. and Stratling, W.H. (1988). The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. *EMBO J.*, **7**, 655-664.

Lowrey, H., Bodine, D. and Nienhuis, A. (1992). Mechanism of DNase I hypersensitivity site formation within the human globin locus control region. *Proc. Natl. Acad. Sci. USA*, **89**, 1143-1147.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.

Magram, J., Chada, K. and Costantini, F. (1985). Developmental regulation of a cloned adult β-globin gene in transgenic mice. *Nature*, **315**, 338-340.

Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987). Regulation of inducible and tissue-specific gene expression. *Science*, **236**, 1237-1245.

Mantovani, R., Superti-Furga, G., Gilman, J. and Ottolenghi, S. (1989). The deletion of the distal CCAAT box region of the $^{A}$γ-globin gene in black HPFH abolishes the binding of the erythroid specific protein NFE3 and of the CCAAT displacement protein. *Nucl. Acids Res.*, **17**, 6681-6691.

Marin, M., Karis, A., Visser, P., Grosveld, F. and Philipsen, S. (1997). Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell*, **89**, 619-628.

Marks, P.A. and Rifkind, R.A. (1978). Erythroleukemic differentiation. *Annu. Rev. Biochem.*, **47**, 419-448.

Marshall, W.F., Fung, J.C. and Sedat, J.W. (1997). Deconstructing the nucleus: global architecture from local interactions. *Curr. Opin. Genet. Dev.*, **7**, 259-263.

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

Martin, P. and Papayannopoulou, T. (1982). HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. *Science*, **216**, 1233-1235.

Mathis, D.J. and Chambon, P. (1981). The 40 early region TATA box is required for accurate *in vitro* initiation of transcription. *Nature*, **290**, 310-315.

McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. and Felsenfeld, G. (1981). A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. *Cell*, 45-55.

McKnight, S. and Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. *Cell*, **46**, 795-805.

Merli, C., Bergstrom, D.E., Cygan, J.A. and Blackman, R.K. (1996). Promoter specificity mediates the independent regulation of neighbouring genes. *Genes Dev.*, **10**, 1260-1270.

Metcalf, D. (1989). The molecular control of cell division, differentiation commitment and maturation in haemopoietic cells. *Nature*, **339**, 27-30.

Mignotte, V., Eleouet, J.F., Raich, N. and Romeo, P.H. (1989). *Cis*- and *trans*-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. *Proc. Natl. Acad. Sci. USA*, **86**, 6548-6552.

Miller, I.J. and Bieker, J.J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol. Cell. Biol.*, **13**, 2776-2786.

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

Mirkovitch, J., Mirault, M.E. and Laemmli, U.K. (1984). Organisation of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*, **39**, 223-232.

Mitchell, P.J. and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, **245**, 371-378.

Mizzen, C.A., Yang, X.J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., Nakatani, Y. and Allis, C.D. (1996). The $TAF_{II}250$ subunit of TFIID has histone acetyltransferase activity. *Cell*, **87**, 1261-1270.

Moi, P., Chan, K., Asunis, I., Cao, A. and 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 the β-globin locus control region. *Proc. Natl. Acad. Sci. USA*, **91**, 9926-9930.

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

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

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucl. Acids Res.*, **9**, 6047-6068.

Muller, H.P., Sogo, J.M. and Schaffner, W. (1989). An enhancer stimulates transcription in *trans* when attached to the promoter via a protein bridge. *Cell*, **58**, 767-777.

Myers, R.M., Tilly, K. and Maniatis, T. (1986). Fine structure genetic analysis of a beta-globin promoter. *Science*, **232**, 613-618.

Ney, P.A., Sorrentino, B.P., McDonagh, K.T. and Nienhuis, A.W. (1990a). 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.

Ney, P.A., Sorrentino, B.P., Lowrey, C.H. and Nienhuis, A.W. (1990b). Inducibility of the HS II enhancer depends on binding of an erythroid-specific nuclear protein. *Nucl. Acids Res.*, **18**, 6011-6017.

Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, **375**, 316-318.

Olsson, I., Gullberg, U., Lantz, M. and Richter, J., (1992). The receptors of regulatory molecules of hematopoiesis. *European journal of haematology*, **48**, 1-9.

Onate, S.A., Tsai, S.Y., Tsai, M.J. and O'Malley, B.W. (1995). Sequence and characterisation of a coactivator for the steroid hormone receptor superfamily. *Science*, **270**, 1354-1357.

O'Neill, L.P. and Turner, B.M. (1995). Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent manner. *EMBO J.*, **14**, 3946-3957.

Orkin, S.H. and Kazazian, H.H. Jr. (1984). The mutation and polymorphism of the human beta-globin gene and its surrounding DNA. *Annu. Rev. Genet.*, **18**, 131.171.

Orkin, S.H., Antonarakis, S.E. and Kazazian, H.H. Jr. (1984). Base substitution at position –88 in a beta-thalassaemic globin gene. Further evidence for the role of distal promoter element ACACCC. *J. Biol. Chem.*, **259**, 8679-8681.

Orkin, S.H. (1986). Disorders of Hemoglobin Synthesis: The Thalassaemias.

Oyake, T., Itoh, K., Motohashi, H., Hayashi, N., Hoshino, H., Nishizawa, M., Yamamoto, M. and 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.

Palmiter, R.D., Sandgren, E.P., Koeller, D.M. and Brinster, R.L. (1993). Distal regulatory elements from the mouse metallothionein locus stimulate gene expression in transgenic mice. *Mol. Cell. Biol.*, **13**, 5266-5275.

Paranjape, S.M., Kamakaka, R.T. and Kadonaga, J.T. (1994). Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annu. Rev. Biochem.*, **63**, 265-297.

Perkins, A.C., Sharpe, A.H. and Orkin, S.H. (1995). Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, **375**, 318-322.

Perkins, A.C., Gaensler, K.M.L. and Orkin, S.H. (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. USA*, **93**, 12267-12271.

Perutz, M.F., Rossman, M.G., Cullis, A.F., Muirhead, H., Will, G. and North, A.C.T. (1960). Structure of haemoglobin. *Nature*, **185**, 416-422.

Perutz, M.F. (1983). Species adaptation in a protein molecule. *Mol. Biol. Evol.*, **1**, 1-28.

Peschle, C., Mavilio, F., Carè, A., Migliaccio, G., Migliaccio, A.R., Salvo, G., Samoggia, P., Petti, S., Guerriero, R., Marinucci, M., Lazzaro, D., Russo, G. and Mastroberardino, G. (1985). Haemoglobin switching in human embryos: asynchrony of $\zeta \rightarrow \alpha$ and $\varepsilon \rightarrow \gamma$-globin switches in primitive and definitive erythropoietic lineage. *Nature*, **313**, 235-238.

Peterson, C.L. (1996). Multiple SWItches to turn on chromatin ?. *Curr. Opin. Genet. Dev.*, **6**, 171-175.

Peterson K.R., Clegg, C.H., Huxley, C., Josephson, B.M., Haugen, H.S., Furukawa, T. and Stamatoyannopoulos. (1993). Transgenic mice containing a 248-kb yeast artificial chromosome carrying the human beta-globin

locus display proper developmental control of human globin genes. *Proc. Natl. Acad. Sci. USA*, **90**, 7593-7597.

Peterson, K.R., Zitnik, G., Huxley, C., Lowrey, C.H., Gnirke, A., Leppig, K.A., Papayannopoulou, T. and Stamatoyannopoulos, G. (1993). Use of yeast artificial chromosomes (YACs) for studying control of gene expression: Correct regulation of the genes of a human beta-globin locus YAC following transfer to mouse erythroleukemia cell lines. *Proc. Natl. Acad. Sci. USA*, **90**, 11207-11211.

Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H. and Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. *Dev.*, **121**, 163-172.

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

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

Pissard, S., M'rad, A., Beuzard, Y. and Romeo, P.H. (1996). A new type of hereditary persistence of fetal haemoglobin (HPFH): HPFH tunisia $\beta^{+}$ (+C-200)$^{G}\gamma$. *British Journal of Haematology*, **95**, 67-72.

Poncz, M., Henthorn, P., Stoeckert, C. and Surrey, S. (1988). Globin gene expression in hereditary persistence of fetal hemoglobin and delta-beta thalassaemia. In Oxford Surveys on Eukaryotic genes, (ed. N. McLean). Oxford University Press. D 5, 163-203.

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

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. *Nature*, **322**, 697-701.

Pugh, B.F. (1996). Mechanisms of transcription complex assembly. *Cell Biology*, **8**, 303-311.

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

Raich, N., Papayannopoulou, T., Stamatoyannopoulos, G. and Enver, T. (1992). Demonstration of a human epsilon-globin gene silencer with studies in transgenic mice. *Blood*, **79**, 861-864.

Raich, N., Clegg, C.H., Grofti, J., Romeo, P.H. and Stamatoyannopoulos, G. (1995). GATA1 and YY1 are developmental repressors of the human epsilon-globin gene. *EMBO J.*, **14**, 801-809.

Reddy, P.M. and Shen, C.K. (1991). Protein-DNA interactions *in vivo* of an erythroid-specific, human beta-globin locus enhancer. *Proc. Natl. Acad. Sci. USA*, **88**, 8676-8680.

Reddy, P.M. and Shen, C.K. (1993). Erythroid differentiation of mouse erythroleukemia cells results in reorganisation of protein-DNA complexes in the mouse beta maj globin promoter but not its distal enhancer. *Mol. Cell. Biol.*, **13**, 1093-1103.

Reitman, M., Lee, E., Westphal, H. and Felsenfeld, G. (1990). Site-independent expression of the chicken $\beta^{A}$-globin gene in transgenic mice. *Nature*, **348**, 749-752.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. *Nature*, **311**, 532-537.

Rochette, J., Craig, J.E. and Thein S.L. (1994). Fetal hemoglobin levels in adults. *Blood reviews*, **8**, 213-224.

Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *TIBS*, **21**, 327-335.

Ronchi, A., Berry, M., Raguz, S., Imam, A., Yannoutsos, N., Ottolenghi, S., Grosveld, F. and Dillon, N. (1996). Role of the duplicated CCAAT box region in γ-globin gene regulation and hereditary persistence of fetal haemoglobin. *EMBO J.*, **15**, 143-149.

Russell, J.E. and Liebhaber, S.A. (1996). The stability of human β-globin mRNA is dependent on structural determinants positioned within its 3' untranslated region. *Blood*, **12**, 5314-5323.

Saha, S., Brickman, J.M., Lehming, N. and Ptashne, M. (1993). New eukaryotic transcriptional repressors. *Nature*, **363**, 648-652.

Schild, C., Claret, F.X., Wahli, W. and Wolffe, A.P.(1993). A nucleosome-dependent static loop potentiates estrogen-regulated transcription from the *Xenopus* vitellogenin B1 promoter *in vitro*. *EMBO J.*, **12**, 423-433.

Schlissel, M.S. and Brown, D.D. (1984). The transcriptional regulation of *Xenopus* 5s RNA genes in chromatin: the role of active stable transcription complexes and histone H1. *Cell*, **37**, 903-913.

Serfling, E., Lubbe,A., Dorsch-Hasler, K. and Schaffner, W. (1985). Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. *EMBO J.*, **4**, 3851-3859.

Sharpe, J., Nonchev, S., Gould, A., Whiting, J. and Krumlauf, R. (1998). Selectivity, sharing and competitive interactions in the regulation of *Hoxb* genes. *EMBO J.*, **17**, 1788-1798.

Shehee, W.R., Oliver, P. and Smithies, O. (1993). Lethal thalassemia after insertional disruption of the mouse major adult beta-globin gene. *Proc. Natl. Acad. Sci. USA*, **90**, 3177-3181.

Shih, D.M., Wall, R.J. and Shapiro, S.G. (1990). Developmentally regulated and erythroid specific expression of the human embryonic ß-globin gene in transgenic mice. *Nucl. Acids Res.*, **18**, 5465-5472.

Shih, D.M., Wall, R.J. and Shapiro, S.G. (1993). A 5' control region of the human epsilon-globin gene is

sufficient for embryonic specificity in transgenic mice. *J. Biol. Chem.*, **268**, 3066-3071.
Shivdasani, R.A. and Orkin, S.H. (1995a). Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc. Natl. Acad. Sci. USA*, **92**, 8690-8694.
Shivdasani, R.A., Rosenwalt, M.F., Zucker-Franklin, D., Jackson, C.W., Hunt, P., Saris, C.J. and Orkin, S.H. (1995b). Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*, **81**, 695-704.
Simon, M.C., Pevny, L., Wiles, M.V., Keller, G., Costantini, F. and Orkin, S.H. (1992). Rescue of erythroid development in gene targeted GATA-1 mouse embryonic stem cells. *Nature Genetics*, **1**, 92-98.
Skow, L.C., Burkhart, B.A., Johnson, F.M., Popp, R.A., Popp, D.M., Goldberg, S.Z., Anderson, W.F., Barnett, L.B. and Lewis, S.E. (1983). A mouse model for beta-thalassaemia. *Cell*, **34**, 1043-1052.
Smale, S.T. and Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell*, **57**, 103-113.
Solomon, M.J. and Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. *Proc. Natl. Acad. Sci. USA*, **82**, 6470-6474.
Southwood, C.M., Downs, K.M. and Bieker, J.J. (1996). Erythroid Krüppel-like factor exhibits an early and sequentially localised pattern of expression during mammalian erythroid ontogeny. *Dev. Dyn.*, **206**, 248-259.
Souza, S.J., Long, M. and Gilbert, W. (1996). Introns and gene evolution. *Genes to Cells*, **1**, 493-505.
Stalder, J., Larsen, A., Engel, J.D., Dolan, M., Groudine, M. and Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNase I. *Cell*, **20**, 451-460.
Stamatoyannopoulos, G. and Nienhuis, A.W. (1987). Hemoglobin switching. *The molecular basis of blood diseases*, Saunder comp. Philadelphia, 66-105.
Stamatoyannopoulos, G. and Nienhuis, A.W. (1989). Hemoglobin switching. Part A: Transcriptional regulation.
Stamatoyannopoulos, G. and Nienhuis, A.W. (1994). Hemoglobin switching. *The molecular basis of blood diseases*, Saunder comp. Philadelphia, 107-155.
Steger, D.J. and Workman, J.L. (1996). Remodelling chromatin structures for transcription: what happens to the histones ? *BioEssays*, **18**, 875-884.
Steward, T.A., Wagner, E.F. and Mintz, B. (1982). Human β-globin gene sequences injected into mouse eggs, retained in adults, and transmitted to progeny. *Science*, **217**, 1046-1048.
Strauss, E.C. and Orkin, S.H. (1992). *In vivo* protein-DNA interactions at hypersensitive site 3 of the human β-globin locus control region. *Proc. Natl. Acad. Sci. USA*, **89**, 5809-5813.
Strick, R. and Laemmli, U.K. (1995). SARs are *cis* DNA elements of chromosome dynamics: Synthesis of a SAR repressor protein. *Cell*, **83**, 1137-1148.
Strouboulis, J., Dillon, N. and Grosveld, F. (1992). Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev.*, **6**, 1857-1864.
Strouboulis, J., Dillon, N. and Grosveld, F. (1992). Efficient joining of large DNA fragments for transgenesis. *Nucl. Acids Res.*, **20**, 6109-6110.
Struhl, K. (1996). Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell*, **84**, 179-182.
Studitsky, V.M., Clark, D.J. and Felsenfeld, G. (1995). Overcoming a nucleosomal barrier to transcription. *Cell*, **83**, 19-27.
Stuve, L.L. and Myers, R.M. (1990). A directly repeated sequence in the beta-globin promoter regulates transcription in murine erythroleukemia cells. *Mol. Cell. Biol.*, **10**, 972-981.
Stuve, L.L. and Myers, R.M. (1993). Identification and characterisation of a β-globin promoter-binding factor from murine erythroleukemia cells. *Mol. Cell. Biol.*, **13**, 4311-4322.
Superti-Furga, G., Barberis, A., Schaffner, G. and Busslinger, M. (1988). The -117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the γ-globin gene. *EMBO J.*, **7**, 3099-3107.
Svaren, J. and Hörz, W. (1996). Regulation of gene expression by nucleosomes. *Genes Dev.*, **6**, 164-170.
Svaren, J. and Hörz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in Yeast. *TIBS*, **22**, 93-97.
Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. and Greaves, D.R. (1989). A dominant control region from the human β-globin locus conferring integration site-independent gene expression. *Nature*, **338**, 352-355.
Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990). Detailed analysis of the site 3 region of the human β-globin dominant control region. *EMBO J.*, **9**, 2169-2178.
Talbot, D. and Grosveld, F. (1991). The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *EMBO J.*, **10**, 1391-1398.
Tang, D.C., Ebb, D., Hardison, R.C. and Rodgers, G..P. (1997). Restoration of the CCAAT box or insertion of the CACCC motif activate δ-globin gene expression. *Blood*, **90**, 421-427.
Taramelli, R., Kioussis, D., Vanin, E. et al. (1986). Gamma delta beta-thalassaemias 1 and 2 are the result of a

100 kbp deletion in the human beta-globin cluster. *Nucl. Acids Res.*, **14**, 7017-7029.
Taylor, I.C., Workman, J.L., Schuetz, T.J. and Kingston, R.E. (1991). Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. *Genes Dev.*, **5**, 1285-1298.
Tazi, J. and Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell*, **60**, 909-920.
Tewari, R., Gillemans, N., Harper, A., Wijgerde, M., Zafarana, G., Drabek, D., Grosveld, F. and Philipsen, S. (1996). The human β-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene. *Dev.*, **122**, 3991-3999.
Tewari, R., Gillemans, N., Wijgerde, M., von Lindern, M., Grosveld, F. and Philipsen, S. (1998). Erythriod Krüppel-like factor (EKLF) is active in primitive and definitive erthroid cells and is required for the function of 5'HS3 of the β-globin locus control region. *In press:EMBO J.*
Thein, S.L. (1993). ß-thalassaemia. *Chapter 6 in "Baillière's Clinical Haematology"*, **6**, 151-175.
Thoma, F., Koller, T. and Klug, A. (1979). Involvement of histone H1 in the organisation of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.*, 403-427.
Thomas, G.H. and Elgin, S.C. (1988). Protein/DNA architecture of the DNase I hypersensitive region of the *Drosophila* hsp26 promoter [published erratum appears in EMBO J 1988 Oct;7(10):3300]. *EMBO J.*, **7**, 2191-2201.
Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J.V., and Leder, P. (1978). Intervening sequences of DNA identified in the structural portion of a mouse ß-globin gene. *Proc. Natl. Acad. Sci.USA*, **75**, 725-729.
Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1985). Erythroid-specific expression of human beta-globin genes in transgenic mice. *EMBO J.*, **4**, 1715-1723.
Treisman, R., Orkin, S.H. and Maniatis, T. (1983). Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature*, **302**, 591-596.
Treisman, R. and Maniatis, T. (1985). Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked DNA. *Nature*, **315**, 73-75.
Trepicchio, W.L., Dyer, M.A. and Baron, M.H. (1993). Developmental regulation of human embryonic β-like globin gene is mediated by synergistic interactions among multiple tissue- and stage-specific elements. *Mol. Cell. Biol.*, **13**, 7457-7468.
Tsai, S.F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G.G. and Orkin, S.H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature*, **339**, 446-451.
Tsai, F.Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W. and Orkin, S.H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, **371**, 221-226.
Tsukiyama, T. and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodelling factor. *Cell*, **83**, 1011-1020.
Tsukiyama, T. Daniel, C., Tamkun, J. and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodelling factor. *Cell*, **83**, 1021-1026.
Tsukiyama, T. and Wu, C. (1997). Chromatin remodelling and transcription. *Curr. Opin. Genet. Dev.*, **7**, 182-191.
Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985). The "beta-like-globin" gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. USA*, **82**, 6384-6388.
Tuan, D.Y., Solomon, W.B., Cavallesco, R., Huang, G. and London, I.M. (1989). Characterisation of a human globin enhancer element. *Prog. Clin. Biol. Res.*
Tuan, D., Kong, S. and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc. Natl. Acad. Sci. USA*, **89**, 11219-11223.
van Assendelft, G.B., Hanscombe, O., Grosveld, F. and Greaves, D.R. (1989). The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell*, **56**, 969-977.
van der Hoeven, F., Zakany, J. and Duboule, D. (1996). Gene transpositions in the *HoxD* complex reveal a hierarchy of regulatory controls. *Cell*, **85**, 1025-1035.
van der Ploeg, , L., Konings, A., Oort, M., Bernini, L. and Flavell, R. (1980). Gamma-beta thalassaemia: deletion of the gamma and delta genes influences beta globin gene expression in man. *Cell*, **283**, 637-642.
Varga-Weisz, P.D. and Becker, P.B. (1995). Transcription factor-mediated chromatin remodelling: mechanisms and models. *FEBS*, **369**, 118-121.
Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M. and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature*, **388**, 598-601.
Verrijzer, C.P., Chen, J.L., Yokomori, K. and Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, **81**, 1115-1125.
Vidali, G., Boffa, L.C., Bradbury, E.M. and Allfrey, V.G. (1978). Butyrate suppression of histone deacetylation leads to accumulation of multi-acetylated forms of histones H3 and H4 and increased DNase I sensitivity of

the associated DNA sequences. *Proc. Natl. Acad. Sci. USA*, **75**, 2239-2243.
Volloch, V. and Housman, D. (1982). Terminal differentiation of murine erythroleukemia cells: Physical stabilisation of end-stage cells. *J. Cell. Biol.*, **93**, 390-394.
Wade, P.A., Pruss, D. and Wolffe, A.P. (1997). Histone acetylation: chromatin in action. *TIBS*, **22**, 128-132.
Wall, L., deBoer, E. and Grosveld, F. (1988). The human β-globin 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev.*, **2**, 1089-1100.
Walters M.c., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. and Martin, D.I.K. (1995). Enhancers increase the probability but not the level of gene expression. *Proc. Natl. Acad. Sci. USA*, **92**, 7125-7129.
Walters, M.C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M. and Martin, D.I. (1996). Transcriptional enhancers act in *cis* to suppress position-effect variegation. *Genes Dev.*, **10**, 185-195.
Warburton, P.E. and Earnshaw, W.C. (1997). Untangling the role of DNA topoisomerase II in mitotic chromosome structure and function. *BioEssays*, **19**, 97-99
Weatherall, D.J. and Clegg, J.B. (1981). The Thalassaemia syndromes. Oxford: Blackwell Scientific.
Webber, A.L., Ingram, R.S., Levorse, J.M. and Tilghman, S.M. (1998). Location of enhancers is essential for the imprinting of *H19* and *Igf2* genes. *Nature*, **391**, 711-715.
Weber, F. and Schaffner, W. (1985). Simian virus 40 enhancer increases RNA polymerase density within the linked gene. *Nature*, **315**, 75-77.
Weintraub, H. and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. *Science*, **193**, 848-856.
Weintraub, H. (1988). Formation of stable transcription complexes as assayed by analysis of individual templates. *Proc. Natl. Acad. Sci. USA*, **85**, 5819-5823.
Weis, L. and Reinberg, D. (1992). Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *Faseb J.*, **6**, 3300-3309.
Weiss, I.M. and Liebhaber, S.A. (1994). Erythriod cell-specific determinants of α-globin mRNA stability. *Mol. Cell Biol*, **14**, 8123 ?
Weiss, M.J., Keller, G. and Orkin, S.H. (1994). Novel insights into erythroid development revealed through *in vitro* differentiation of GATA-1 embryonic stem cells. *Genes Dev.*, **8**, 1184-1197.
Wijgerde, M., Grosveld, F. and Fraser, P. (1995). Transcription complex stability and chromatin dynamics *in vivo*. *Nature*, **377**, 209-213.
Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. and Fraser, P. (1996). The role of EKLF in human β-globin gene competition. *Genes Dev.*, **10**, 2894-2902.
Wilson, C., Bellen, H.J. and Gehring, W.J. (1990). Position effects on eukaryotic gene expression. *Annu. Rev. Cell. Biol.*, **6**, 679-714.
Wilson, C.J., Chao, D.M., Imbalzano, A.N., Schnitzler, G.R., Kingston, R.E. and Young, R.A. (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodelling. *Cell*, **84**, 235-244.
Winoto, A. and Baltimore, D. (1989). Alpha beta lineage-specific expression of the alpha T cell receptor gene by nearly silencers. *Cell*, **59**, 649-655.
Wolffe, A.P. and Pruss D. (1996). Deviant nucleosomes: the functional specialisation of chromatin. *TIG*, **12**, 58-62.
Wood, W.G. (1993). Increased HbF in adult life. Chapter 7 in "Baillière's clinical Haematology", Baillière Tindall, vol . 6, no. 1, March 1993.
Wright, S., deBoer, E., Grosveld, F.G., Flavell, R.A. (1983). Regulated expression of the human ß-globin gene family in murine erythroleukemia cells. *Nature*, **305**, 333-336.
Wright , S., Rosenthal, A., Favell, R., and Grosveld, F. (1984). DNA sequences required for regulated expression of ß-globin genes in murine erythroleukemia cells. *Cell*, **38**, 265-273.
Xiao, J.H., Davidson, I., Macchi, M., Rosales, R., Vigneron, M., Staub, A. and Chambon, P. (1987). *In vitro* binding of several cell-specific and ubiquitous nuclear proteins to the GT-I motif of the SV40 enhancer. *Genes Dev.*, **1**, 794-807.
Yang, W.M., Inouye, C., Zeng, Y., Bearss, D. and Seto, E. (1996). Transcriptional repression by YY1 is mediated by interaction with a mammalian homologue of the yeast global regulator RPD3. *Proc. Natl. Acad. Sci. USA*, **93**, 12845-12850.
Yang, X.J., Ogryzko, V.V., Nishikawa, J.I., Howard, B.H. and Nakatani, Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature*, **382**, 319-324.
Yu, C.Y., Chen, J., Lin, L.I. et al. (1990). Cell type-specific protein-DNA interactions in the human ζ-globin upstream promoter region: displacement of Sp1 by erythroid cell-specific factor NF-E1. *Mol. Cell. Biol.*, **10**, 282-294.
Zhao, k., Kas, E., Gonzalez, E. and Laemmli, U.K. (1993). SAR-dependent mobilisation of histone H1 by HMG-I/Y *in vitro*: HMG-I/Y is enriched in H1-depleted chromatin. *EMBO J.*, **12**, 3237-3247.

# Chapter 2

Transcription complex stability and chromatin dynamics in vivo

*Wijgerde, M., Grosveld, F. and Fraser P.*
*Nature 377, 209-213, (1995)*

# Transcription complex stability and chromatin dynamics *in vivo*

**Mark Wijgerde*, Frank Grosveld*† & Peter Fraser***

* MGC Department of Cell Biology and Genetics, Erasmus University, 3015 GE Rotterdam, The Netherlands
† Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

**Distant regulatory sequences affect transcription through long-range chromatin interactions. Visualization of transcriptional activity of genes that compete for distant elements, using the globin locus as a model, has revealed the dynamics of chromatin interactions *in vivo*. Multiple genes appear to be transcribed alternately rather than at the same time to generate several messenger RNAs in one cell. The regulator may stably complex with one gene at a time and switch back and forth between genes in a flip-flop mechanism.**

MANY genes are dependent for expression on the presence of distant regulatory elements which may be tens of thousands of base pairs away. The five human β-globin genes are arranged in the order of their developmental expression (ε-Gγ-Aγ-δ-β)[1] and all are dependent on the locus control region (LCR) for high-level, position-independent expression[2-4]. The LCR is located over 50 kilobases (kb) upstream of the β-globin gene and consists of five DNase I-hypersensitive sites (HS)[3,5-7]. The most important sites in terms of transcriptional activation are HS 2–4 (refs 8, 9 and J. Ellis *et al.*, manuscript submitted), each having a core region of 200–300 base pairs (bp)[10-14].

Transcriptional competition between genes[15,16] is important *in vivo* in determining the pattern of globin gene expression during development[17-20] (N. Dillon *et al.*, manuscript submitted). All of the genes are in polar competition for the activating function of the LCR at all stages of development, with proximal genes having an advantage over distal genes, suggesting that the LCR interacts directly with the gene(s) by a looping mechanism.

The distal β-globin gene is transcriptionally competent at all stages of development, but its expression is suppressed in embryonic erythroid cells by the LCR proximal ε- and γ-globin genes. Expression of the β-gene occurs only after silencing of the ε- and γ-genes, presumably by multiple silencing elements in the sequences immediately flanking the genes[21-23]. In contrast to models proposing that different HS regions of the LCR interact with different genes at the same time[24], we have proposed that the LCR HS elements interact to act together or form a holocomplex which then interacts with proximal transcriptionally competent genes via looping (ref. 9 and J. Ellis *et al.*, submitted). The LCR holocomplex model would explain why the activity of the HS is additive and why each of the genes requires all of the HS for full expression. Most important, it would explain the fact that there is balanced competition between the genes for LCR function. A key question in validating this model is whether the LCR is limited to activating only a single gene at a time or, as others suggest, the LCR splits its function[24] and activates several genes simultaneously[25,26]. We have addressed this question by *in situ* hybridization analysis of the erythroid tissues derived from a transgenic mouse line carrying a single copy of the complete human β-globin locus. The results show that the LCR activates only one gene at a time, indicating that the LCR–globin-gene interaction is monogenic. Furthermore, these interactions are not static, but dynamic, with transcription interactions forming, breaking and reforming in a type of flip-flop mechanism. Finally, the average stability of the LCR–gene interactions was estimated to be of the order of 15–80 min.

## Globin primary transcripts

Analysis of γ- and β-globin steady-state messenger RNA levels during switching in the early stages of transgenic fetal liver erythropoiesis (11.5–13.5 d) show that both are detectable in total fetal liver RNA[27], with γ-gene expression decreasing and β-globin gene expression increasing. Both γ- and β-globin polypeptides occur in most cells of the transgenic fetal liver[9,28], indicating that commitment to γ- or β-gene expression does not occur before the onset of globin transcription. The detection of globin mRNA or polypeptides are not reliable indicators of concurrent transcriptional activity in developing cells owing to the long half-lives of such molecules. Alternatively, primary transcripts are both temporally and spatially associated with the transcriptional event and have short half-lives due to rapid splicing into mature mRNA. The half-life of the mouse β-major globin primary transcript has been calculated to be <5 min in mouse erythroleukaemia cells[29]. Detection of primary transcripts *in situ* is therefore an accurate indicator of ongoing or very recent transcription. If transcription complexes last longer than the half-life of the primary transcript, two types of signals may occur: single-gene transcription signals or multigene signals. Whereas the explanation for the single-gene signal is obvious, the multigene signal could result from either of two events, namely concurrent transcription from two genes or a recent switch causing a signal overlap between transcription from one

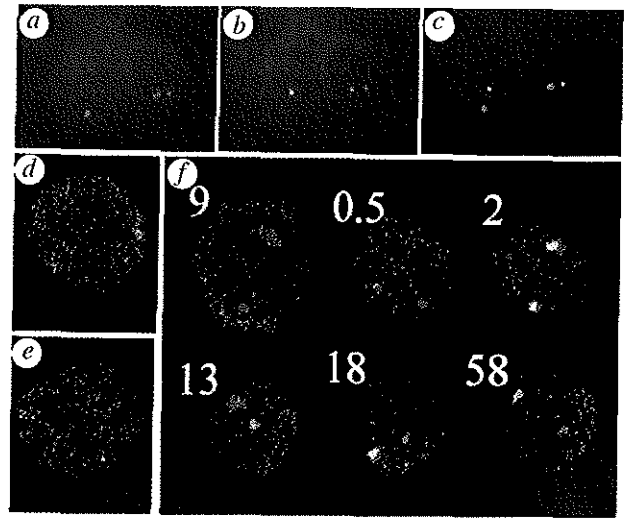



FIG. 1 *In situ* hybridization of transgenic and non-transgenic mouse fetal liver cells. *a–c*, Hybridization of homozygous 12.5-day mouse fetal liver cells (line 72)[27] containing two single human β-globin loci; *a* shows the red signal only (Texas red, γ), *b* the green signal only (FITC, β), and *c* the overlay of both after laser confocal microscopy. *d*, *e*, same as *c*, but in *d* cells have first been treated with RNase; *e* shows results from a non-transgenic fetal liver. *f*, Quantification of the different types of transgenic mouse 12.5-day fetal liver cells. Red (Texas red) represents the γ signal, green (FITC) the β signal, and yellow the combination of both. Scoring was by epifluorescence on a minimum of 400 cells per liver sample and at least two livers. The frequency of occurrence of each cell type is shown as a percentage.

METHODS. 12.5-day fetal livers were disrupted into PBS and the cells fixed onto a poly-L-lysine coated slide in 4% formaldehyde/5% acetic acid for 20 min at room temp.[35]. Cells were subsequently washed 3 times for 10 min in PBS and stored in 70% ethanol at −20 °C. Slides were pretreated for hybridization by a 0.01% pepsin digestion (5 min, 37 °C) in 0.01 M HCl, followed by a short wash in water and a 5-min fixation in 3.7% formaldehyde at room temperature, then washed in PBS, dehydrated in 70, 90 and 100% ethanol steps and air-dried. The hybridization mixture was applied (12 μl per 24 × 24 mm coverslip) and incubated at 37 °C in a moisturized chamber for 12 h. The hybridization mixture contained 0.5 ng $\mu l^{-1}$ of each of three or four oligonucleotides (50 nucleotides long) containing a DNP, a digoxygenin or a biotin side chain in the middle and on the 5′ and the 3′ end of the oligonucleotide (Eurogentec, Belgium) in 25% formamide, 2× SSC, salmon sperm DNA (200 ng $\mu l^{-1}$), 5× Denhardts, 1 mM EDTA and 50 mM sodium phospate, pH 7.0. The oligonucleotide sequences were derived from the first and second introns of the globin genes and spaced more than 25 nucleotides apart. The coverslip was removed by dipping in 2× SSC and the cells were washed three times in 2× SSC at 37 °C, followed by a 5-min wash in 0.1 M Tris, 0.15 M NaCl, 0.05% Tween 20. Antibody detection of the labels was essentially as described by Dirks *et al.*[35], with three or four amplification steps. Mounting was in DAPI/DABCO:Vectashield (1:1) in glycerol (90%) and stored at 4 °C in the dark. Fluorescence was detected by epifluorescence/CCD or laser confocal microscopy. RNase treatment was in 0.1 M Tris, 0.15 M NaCl, 10 μg $ml^{-1}$ RNase A for 5 min at room temperature.

gene and the decay of the pre-mRNA (processing) from the other gene.

We have used *in situ* hybridization with gene-specific intron probes to localize human globin primary transcripts in nuclei of individual embryonic and fetal transgenic erythroid cells. The signals appear in the nucleus as fluorescent foci at the location of an actively transcribing gene[30,31]. The foci do not appear in RNase-pretreated cells or in non-globin-expressing transgenic cells or non-transgenic cells (Fig. 1*d*, *e*), indicating that the probes specifically detect an intron-containing RNA molecule (the primary transcript). Probe penetration is almost complete as >97% of the active mouse α-globin genes can be detected (not shown). Heterozygous cells show single foci, whereas homozygous cells show two foci, demonstrating detection of each locus (not shown). To determine whether the LCR can activate the γ- and β-globin genes simultaneously, we performed *in situ* hybridizations with gene-specific intron probes for γ- and β-globin primary transcripts in homozygous day-12 transgenic mouse fetal liver cells (Fig. 1*a*–*c*). The cell on the right contains primary transcript signals for both γ- (red) and β-globin (green) from each chromosome, suggesting that the LCR can activate both genes simultaneously from a single chromosome, as would be expected for co-transcription. Alternatively, the double signal could be due to an overlap between transcription of one gene and the decay of the pre-mRNA signal from the other gene, as would be expected for single-gene transcription. However, only the latter explanation would also fit the type of cell on the left, which is transcribing only γ-globin from one chromosome and only β-globin from the other. This suggests that the LCR activation mechanism is mono-gene-specific. Considering that these nuclei contain the transacting factors required for both γ- and β-globin transcription, co-transcription would be unlikely if the γ and β-only cell (left) were to occur frequently, given the high probe penetration (>97%). We therefore quantified the different cell types.

## A dynamic mechanism

Day-12 homozygous transgenic fetal liver cells show all possible primary transcript signal combinations of γ- and β-globin (Fig. 1*f*): 58% of globin-transcribing nuclei have only β-globin transcription on both chromosomes and 9% have only γ-globin; the remaining one third of the erythroid cells contain combinations of γ- and β-globin transcription and are therefore involved in the switching process. However, fewer than half (34%) of the chromosomes in these switching cells have both γ- and β-globin signals on the same chromosome. The majority have single gene signals only, suggesting that the LCR interaction is largely mono-gene-specific. Of particular interest is the fact that in >90% of the switching cells, the two chromosomes are responding differently to the same transacting-factor environment. This suggests a dynamic, continuously changing system in which the individual loci respond stochastically to changes in the factor environment.

## Flip-flop

Two forms of such a dynamic process can be envisaged. Either the switch in interaction is progressive from γ to β, or it switches back and forth between the genes in a kind of flip-flop mechanism. γ- and β-globin signals on the same chromosome would then be indicative of a recent switch producing a (brief) period of overlap in which the decaying 'old' primary transcripts are detected in the presence of the newly synthesized transcripts. The length of this overlap period would depend on the half-life of the primary transcript. The progressive switching mechanism

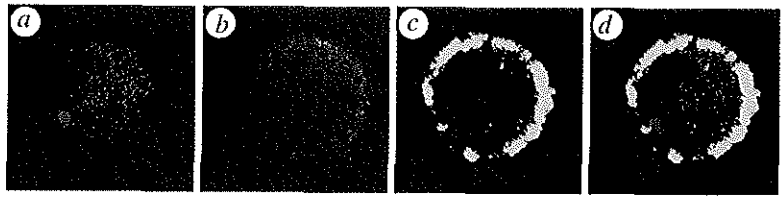



FIG. 2 Detection of γ-globin (Texas red) and β-globin (FITC, green) precursor mRNA in the nucleus and β-globin mRNA in the cytoplasm (Cy3, yellow). The fetal liver was obtained from a transgenic 12.5-day embryo heterozygous for the human β-globin locus. β-globin mRNA was visualized by indirect immunofluorescence with Cy3 of DNP-labelled oligonucleotides. Cy3 emission is also red but at a different wavelength from Texas red and was therefore recorded separately and assigned an artificial yellow. *a*, Texas red signal (γ pre-mRNA); *b*, FITC green signal (β pre-mRNA)—this image was digitally enhanced to demonstate the lack of a β transcript, so some bleed-through from the cytoplasmic Cy3 signal appears as a high background; *c*, Cy3 signal (β mRNA). This is a single optical section of the cell from the confocal microscope (0.5 microns); *d* is a composite of *a*, *b* and *c*.

would predict that heterozygous γ-gene-transcribing cells should not have β-mRNA in their cytoplasm, whereas the flip-flop mechanism would result in a proportion of γ (but not β) transcribing nuclei, which contain β mRNA in their cytoplasm.

We therefore added a third probe which specifically recognizes the human β-globin mRNA (exon probe). Heterozygous 12.5-day transgenic fetal liver cells were used, needing only a single chromosome to switch. Figure 2 shows one of many heterozygous cells probed for the γ- and β-globin primary transcripts and cytoplasmic β-globin mRNA. The γ transcription signal is very strong (red in Fig. 2*a*) whereas the β-gene signal (green in Fig. 2*b*) is absent, indicating that the γ gene is currently being transcribed, but the β gene is not. The cytoplasm, however, shows an accumulation of β mRNA, indicating previous transcriptional activity of the β-gene in that cell (Fig. 2*c*). In fact, 40% of the erythroid cells that are transcribing the γ gene, but not the β gene, contain β mRNA. In other words, there is a substantial number of cells that have transcribed β but switched back to γ, an observation that strongly supports a dynamic flip-flop between the genes. Many of the chromosomes with both γ- and β-globin primary transcript signals (Fig. 2) would therefore be the result of flip-flop rather than co-transcription. Thus, the decay time of our signal after a switch becomes important in determining whether most or all of the overlap we observe is due to flip-flop. The 15S mouse β-major globin primary transcript was shown to reach its steady-state level in MEL cells within 5 min[29]. This indicates that the half-life of the completely intact precursor RNA is considerably shorter as it takes ~5 half-lives to reach 97.5% of steady-state level. However, the target of our *in situ* experiments is not only the intact 15S globin primary transcript, but also any other intron-containing RNA molecules such as partially transcribed, partially spliced and excised intron RNAs. We therefore measured the half-life of the intron-containing RNA molecules for mouse β-major globin and human γ- and β-globin in day-12 transgenic fetal liver cells. The results indicate that all three have similar kinetics, reaching steady-state levels simultaneously (Fig. 3). The half-life of the intron is 4–5 min, in agreement with Curtis *et al.*[29]. Assuming that we can still detect the *in situ* signal after 2–3 half-lives (see discussion), it suggests that the overlap period is 10–15 min.

## Replication and switching

One possibility to explain the occurrence of multigene signals when only a single gene is active at any time could be that it is the result of replication of a locus. Homozygous replicating cells (in late S, G2 and M) contain four globin loci in the nucleus and hence a γβ co-transcription signal could be the result of the γ gene being transcribed from one of the replicated loci, whereas β could be transcribed from its replicated chromatid. This would result in the percentage of γβ signals being substantially higher in G2 than in G1 cells. We therefore sorted 12.5-day fetal liver into populations of G1 and G2 cells[32] and repeated the hybridizations. Analysis of the G1 and G2 cells shows a distribution and percentage of γβ signals similar to the total population, with no significant difference between the two purified populations. We conclude that the switch between genes can take place in G1—that is, DNA replication is not required, implying that the LCR–gene interactions are dynamic rather than static. As expected[33,34], no transcription signals are observed when the cells are in mitosis.

## Co-expression and switching

To extend these results, we characterized the transcriptional status of individual cells at different times of development,

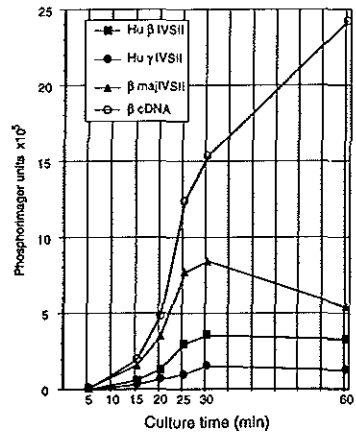



FIG. 3 Incorporation of $^{32}P$ into globin intron and mRNA sequences. 12.5-day fetal liver cells from *heterozygous* line 72 were cultured in the presence of $^{32}P$ for the times indicated, essentially as described[29]. Half-lives were calculated by comparing the difference in the rates of incorporation between intron and exon sequences. The 4–5-min half-life fits well with the fact that 5 half-lives are required to reach 97.5% of the steady state level.
METHODS. RNA was prepared for $1.5 \times 10^7$ cells per time point and hybridized to 5 μg immobilized denatured plasmid DNA as described[35]. Results were quantified by phosphorimage analysis. Plasmids contained mouse β-major intron II (*BamHI*/*PstI*), human β intron II (*BamHI*/*SspI*), human γ intron II (*BamHI*/*SacI*) and partial human β-globin cDNA; signals were normalized for relative probe length.

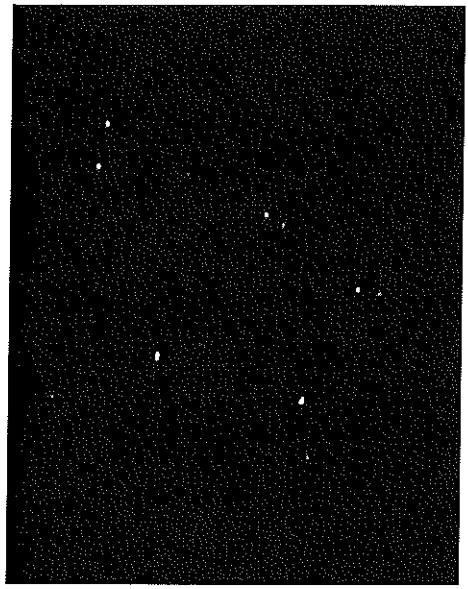


FIG. 4 Simultaneous detection of ε (FITC, green) and γ (Texas red) transcription in 10-day embryonic blood. Cells were obtained from a line-72 embryo homozygous for the human globin locus[27].

namely day-10 embryonic blood and day-11 fetal liver. During transgenic embryonic erythropoiesis γ- and ε-globin expression occurs simultaneously over a period of at least 3 days (days 8–10 of gestation) and nearly all cells contain both ε- and γ-globin polypeptides[28]. We have previously determined the level of expression of the γ- and ε-globin genes when present as single genes linked to the LCR and compared this to the level of expression when both are linked together to the LCR. That comparison suggested that the γ and ε genes are also in competition for the LCR (P.F., unpublished results), presumably through a looping mechanism (N. Dillon *et al.*, submitted). The circulating embryonic erythrocytes are nucleated and transcriptionally active, and hence were used to determine whether flip-flop could also be occurring between the ε- and γ-globin genes by probing day-10 embryonic erythrocytes with gene-specific intron probes for ε and γ (Fig. 4). The results demonstrate (like γ and β in the fetal liver) a heterogeneity in transcriptional signals, including chromosomes with overlapped signals (Figs 4 and 5). We conclude that the LCR also productively interacts with only one gene at a time in peripheral embryonic blood cells and that ε and γ are co-expressed via a flip-flop mechanism. The higher incidence of overlapped signals on individual chromosomes suggests that the LCR–ε complex is less stable than those of γ and β with the LCR (see discussion).

The analysis of day-11 fetal liver shows that the switching process is a continuously changing system, reflected at the level of the individual cell (Fig. 5). When compared to the day-12 results, there are more γγ- and fewer ββ-transcribing cells (27 versus 9% and 35 versus 58%); accordingly, significantly more cells are observed that transcribe γ only from locus and γ and β from the other locus (7 versus <1%). The number of cells transcribing γ only on one chromosome and β only on the other chromosome is very similar (14 versus 13%). These results indicate that the change in the composition of transcription factors during the switch is gradual and slowly changes the affinity between the LCR and the genes towards the adult type LCR/β-globin gene interaction.

## Discussion

We have demonstrated the dynamic nature of the interactions between the human β-globin genes and the β-globin LCR throughout development and shown that the LCR–gene interactions are highly stable. Our data can all be explained by a mechanism in which the LCR activates only a single gene in the locus at any given time. The LCR, although made up of a series of hypersensitive sites, could act as a functional unit or holocomplex that activates multiple genes from the same chromosome via a stochastic mechanism of dynamic rather than static interactions (Fig. 6). Although a low level of co-transcription cannot be excluded such a mechanism cannot account for our observations. Dillon *et al.* (manuscript submitted) have used steady-state measurements of the transcriptional output from mutated loci to demonstrate that the genes compete in a polar manner for the activating function of the LCR. They find that the relative distance of the genes from the LCR affects competition and conclude that the genes are activated by direct interaction with the LCR. Thus two studies, which use entirely different approaches, provide strong evidence of direct interaction between the LCR and single genes by a dynamic looping mechanism of transcriptional activation. It does not show whether some process spreading in a linear fashion along the chromatin is involved in the first activation of the locus. The balance of, γ- versus β-gene transcription would gradually change during a

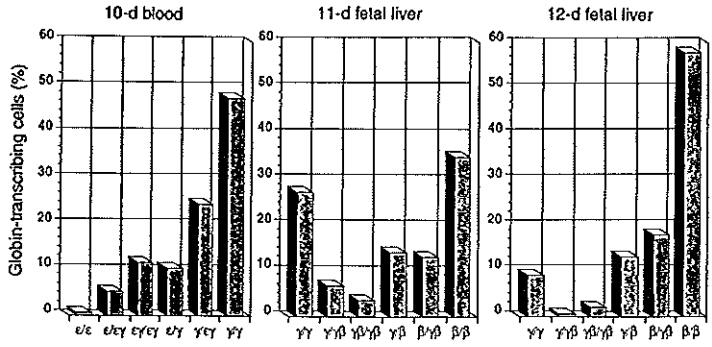



FIG. 5 Bar diagram of the percentages of transcriptional cell types observed at different times of development in transgenic cells homozygous for the human globin locus. A single transcription signal at one of the loci is indicated with a single letter, double signals with two letters. For example γ/γβ is a cell type with a γ signal only on one chromosome and signals for γ and β on the other chromosome. The number of cell types in the population are given as a percentage.

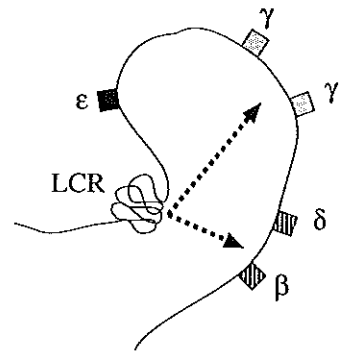



FIG. 6 A model of the interactions of the locus control region (LCR) with the different globin genes through a stochastic looping mechanism. The LCR is indicated as a squiggly line to indicate that the different regions that are hypersensitive to DNase I of the LCR could act together or even form a holocomplex to establish an interaction of the LCR with one of the genes.

switching period as the factor composition changes in the population.

## Flip-flop time and complex stability

By scoring the different transcriptional chromosome types in the erythroid nuclei (Fig. 5), it is possible to estimate the flip-flop time, or the average time required for the LCR–gene interaction to change from γ to β (or vice versa), based on the half-life of the intron-containing RNAs and the percentages of different transcriptional states. The half-life determines the decay rate of the signal and thus the overlap time, that is, the length of time both γ- and β-signals are detectable on the same chromosome after a change in LCR–gene interaction. As the cell population is random (so the chromosomes are not synchronized in terms of the γ-to-β switch), the percentage of chromosomes showing a double signal at any particular instant indicates the fraction of the total time in which a chromosome is in the overlapped state. By extrapolation, we can determine the average length of time that a gene is transcribed—that is, the time between flip-flop events. For example, the frequency of overlapped β signals in the population at 12 days is 13% of the total number of β signals. Therefore, on average, 13% of the time in which an individual chromosome transcribes the β gene it will be overlapped by a γ signal. The overlap time can be estimated as 10–15 min on the basis of the intron half-life of 4–5 min and the fact that we can barely detect the precursor RNA when we use one (instead of three or four) oligonucleotides (M.W., unpublished results). As 13% of the β signal is overlap, the total β gene transcription time would be ~80 min. (As overlap occurs at the beginning and end of a transcription period, the time in overlap is 2 × 10–15 min. The total transcription time is therefore 100/13 × 12 min minus one overlap of 12 min.) Correspondingly, the percentage of overlapped γ signals is 42%, making the average LCR/γ gene time between flip-flops ~16 min. When the results are compared for ε, γ and β from days 10–12, several trends become apparent. First, at 10 days, the ε signal is overlapped with a γ signal 75% of the time, which suggests that the lifetime of the ε-LCR complex is short (about 4 min) and flip-flop occurs more readily. The γ signal is present more often than ε at 10 days and is overlapped only 29% of the time, suggesting that the γ-LCR complex is more stable, with ~30 min between flip-flops. The average stability of the γ-LCR complex decreases from 10 to 12 days from 30 to 16 min, whereas β increases from 45 to 80 min from 11 to 12 days. Clearly the time between flip-flops decreases for γ but increases for β. Obviously interaction could consist of rapid associations and dissociations of the LCR with a given gene, rather than one stable association. Whatever the mechanism, our data suggest that the continuous presence of the regulatory elements at a gene are required for multiple rounds of initiation of transcription of that gene, rather than being dispensable after the activation of transcription.

In summary, our technique has enabled us to analyse single genes in single cells and allows us to make the conclusion that γ- and β-globin genes are transcribed alternately rather than concurrently. The ability to visualize specifically the transcriptional activity of individual genes in the nucleus provides a powerful tool for studying developmental transcriptional regulation and chromatin dynamics and should provide a more accurate picture of transcriptional regulation compared with conventional *in situ* analysis. □

Received 20 February; accepted 4 August 1995.

1. Collins, F. S. & Weissman, S. M. *Prog. nucl. Acid Res. molec. Biol.* **31**, 315–462 (1984).
2. Kioussis, D. *et al. Nature* **308**, 662–666 (1983).
3. Grosveld, F., Blom van Assendelft, G., Greaves, D. R. & Kollias, G. *Cell* **51**, 975–985 (1987).
4. Driscoll, M. C., Dobkin, C. S. & Alter, B. P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7470–7474 (1989).
5. Tuan, D., Solomon, W., Li, Q. & London, I. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6384–6388 (1985).
6. Forrester, W. C., Novak, U., Gelinas, R. & Groudine, M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5439–5443 (1989).
7. Ryan, T. M. *et al. Genes Dev.* **3**, 314–323 (1989).
8. Fraser, P., Hurst, J., Collis, P. & Grosveld, F. *Nucleic Acids Res.* **18**, 3503–3508 (1990).
9. Fraser, P., Pruzina, S., Antoniou, M. & Grosveld, F. *Genes Dev.* **7**, 106–113 (1993).
10. Philipsen, S., Talbot, D., Fraser, P. & Grosveld, F. *EMBO J.* **9**, 2159–2167 (1990).
11. Talbot, D., Philipsen, S., Fraser, P. & Grosveld, F. *EMBO J.* **9**, 2169–2178 (1990).
12. Ney, P. A., Sorrentino, B. P., McDonagh, K. T. & Nienhuis, A. W. *Genes Dev.* **4**, 993–1006 (1990).
13. Pruzina, S. *et al. Nucleic Acids Res.* **19**, 1413–1419 (1991).
14. Strauss, E. & Orkin, S. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5809–5813 (1992).
15. de Villiers, J., Olson, C., Banerji, J. & Schaffner, W. *Cold Spring Harbor Symp. quant. Biol.* **47**, 911–919 (1982).
16. Wasylyk, B., Wasylyk, C., Augereau, P. & Chambon, P. *Cell* **32**, 503–514 (1983).
17. Choi, O. & Engel, J. *Cell* **55**, 17–26 (1988).
18. Behringer, R. R. *et al. Genes Dev.* **4**, 380–389 (1990).
19. Enver, T. *et al. Nature* **344**, 309–313 (1990).
20. Hanscombe, O. *et al. Genes Dev.* **5**, 1387–1394 (1991).
21. Dillon, N. & Grosveld, F. *Nature* **350**, 252–254 (1990).
22. Lindenbaum, M. & Grosveld, F. *Genes Dev.* **4**, 2075–2085 (1990).
23. Raich, N. *et al. Science* **250**, 1147–1149 (1990).
24. Engel, D. *Trends genet. Sci.* **9**, 304–309 (1993).
25. Bresnick, E. H. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1314–1317 (1994).
26. Furukawa, T., Zitnik, G., Leppig, K., Papayannopoulou, T. & Stamatoyannopoulos, G. *Blood* **83**, 1412–1419 (1994).
27. Strouboulis, J., Dillon, N. & Grosveld, F. *Genes Dev.* **6**, 1857–1864 (1992).
28. Strouboulis, J. thesis, Open Univ., UK (1994).
29. Curtis, P., Mantei, N. & Weissmann, C. *Cold Spring Harbor Symp. quant. Biol.* **42**, 971–984 (1977).
30. Lawrence, J. B., Singer, R. H. & Marselle, L. M. *Cell* **57**, 493–502 (1989).
31. Zhang, G., Taneja, K. L., Singer, R. H. & Green, M. R. *Nature* **372**, 22–29 (1994).
32. Arndt-Jovin, D. & Jovin, T. J. *hist. Cytochem.* **25**, 585–589 (1985).
33. Prescott, D. & Bender, M. *Expl Cell Res.* **26**, 260–268 (1962).
34. Terasima, T. & Tolmach, L. *Expl Cell Res.* **30**, 344–362 (1963).
35. Dirks, R., van de Rijke, F., Fujishita, S., van der Ploeg, M. & Raap, A. *J. Cell Sci.* **104**, 1187–1197 (1993).
36. Fraser, P. & Curtis, P. J. *Genes Dev.* **1**, 885–861 (1987).

ACKNOWLEDGEMENTS. We thank M. Antoniou for his help at the start of this project and with the design of the probes; K. Weiss, G. Padella and A. Lamond for assistance with the EMBL confocal microscope; N. Dillon for suggestions; T. Verkerk for cell sorting; L. Braam for animal care; T. Trimborn for assistance; and Eurogentec (Belgium) for help with the probes. M.W. is supported by the NWO (The Netherlands). This work was supported by the NWO (N.U), the MRC (UK) and the Howard Hughes Foundation (USA).

# Chapter 3

The role of EKLF in human β-globine gene competition

*Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. and Fraser, P.*
*Genes & Development 10, 2894-2902, (1996)*

# The role of EKLF in human β-globin gene competition

**Mark Wijgerde, Joost Gribnau, Tolleiv Trimborn, Beatriz Nuez, Sjaak Philipsen, Frank Grosveld, and Peter Fraser[1]**

Medical Genetics Center Department of Cell Biology and Genetics, Erasmus University Rotterdam, 3000 DR, Rotterdam, The Netherlands

**We have investigated the role of erythroid Kruppel-like factor (EKLF) in expression of the human β-globin genes in compound EKLF knockout/human β-locus transgenic mice. EKLF affects only the adult mouse β-globin genes in homozygous knockout mice; heterozygous mice are unaffected. Here we show that EKLF knockout mice express the human ε and γ-globin genes normally in embryonic red cells. However, fetal liver erythropoiesis, which is marked by a period of γ- and β-gene competition in which the genes are alternately transcribed, exhibits an altered ratio of γ- to β-gene transcription. EKLF heterozygous fetal livers display a decrease in the number of transcriptionally active β genes with a reciprocal increase in the number of transcriptionally active γ genes. β-gene transcription is absent in homozygous knockout fetuses with coincident changes in chromatin structure at the β promoter. There is a further increase in the number of transcriptionally active γ genes and accompanying γ gene promoter chromatin alterations. These results indicate that EKLF plays a major role in γ- and β-gene competition and suggest that EKLF is important in stabilizing the interaction between the Locus Control Region and the β-globin gene. In addition, these findings provide further evidence that developmental modulation of globin gene expression within individual cells is accomplished by altering the frequency and/or duration of transcriptional periods of a gene rather than changing the rate of transcription.**

[*Key Words*: Erythroid Kruppel-like factor; locus control region; gene competition; β-globin; transcription]

Received July 23, 1996; revised version accepted October 8, 1996.

---

The human β-globin locus contains five functional erythroid-specific genes arranged in the order of their developmental expression (5′-ε–$^{G}$γ–$^{A}$γ–δ–β-3′) (for review, see Collins and Weissman 1984). The sequences required for correct developmental globin gene regulation have been the target of intensive studies in transgenic mice. Early studies in which individual globin genes were introduced were hampered by lack of expression or low non-copy number-dependent expression, suggesting that expression was dependent on the position of integration in the mouse genome. Although made difficult by these position effects the results suggested that the individual genes contained information neccessary for proper developmental regulation. The inclusion in globin gene constructs of the locus control region (LCR), which consists of 5 DNAse I hypersensitive sites (HS:1–5) located upstream of the cluster, permitted reliable position-independent, copy number-dependent expression at levels equivalent to the endogenous mouse globin genes (Grosveld et al. 1987). Subsequent experiments with individual globin genes linked to the full LCR have shown that the ε- and γ-globin genes are regulated autonomously (Raich et al. 1990; Dillon and Grosveld 1991). The ε gene is restricted to embryonic red cells and the γ gene is expressed in both embryonic- and fetal-derived red cells until it is silenced autonomously around day 16. The adult β gene, on the other hand, is expressed aberrantly at early stages when linked to the LCR and is regulated properly only when another gene is placed in *cis* between it and the LCR, suggesting that the β gene is regulated competitively (Hanscombe et al. 1991). Transgene experiments in which γ and β gene order with respect to the LCR has been varied have demonstrated a developmental expression pattern altering effect (Hanscombe et al. 1991; Peterson and Stamatoyannopoulos 1993; Dillon et al. 1995; N. Dillon and F. Grosveld, in prep.). When a marked β gene is placed in the ε position in the context of the full locus it is expressed throughout development as expected, but, remarkably, it also completely suppresses the fetal and adult expression from the downstream γ and β genes in the locus (Dillon et al. 1995; N. Dillon and F. Grosveld, in prep.).

These data have led us to propose a model for LCR-driven β-globin gene expression in which the individual HS of the LCR act together (or form a holocomplex) and interact directly with an individual globin gene. In this model the complete LCR plays a pivotal role in gene competition. This is supported by the fact that no individual HS of the LCR provides full expression to a linked globin gene (Fraser et al. 1990, 1993), and that all sites are required in the context of the full locus for high-level,

[1]Corresponding author.

position-independent expression (Milot et al. 1996). The competitive advantage of an individual gene is achieved through proximity to the LCR and the specific *trans*-acting factor environment of a particular developmental stage. We propose that these parameters are the basic determinants that control the frequency and duration of direct LCR/gene interactions, which in turn control transcription. We have shown through in situ primary transcript detection that coexpression of multiple genes from a single locus involves alternate rather than cotranscription (Wijgerde et al. 1995). These results indicate that dynamic LCR–gene interactions are the mechanisms of gene competition and coexpression. Obviously, an important question is, which proteins mediate complex formation and stability?

Erythroid Kruppel-like factor (EKLF) is a protein with three zinc fingers that show homology to members of the Kruppel-like family of nuclear proteins such as TFIIIA and Sp1 (Miller and Bieker 1993). EKLF is largely restricted to erythroid cells and binds CACC-box sequences (CCACACCCT) (Donze et al. 1995) frequently found in erythroid specific gene promoters and duplicated in the adult mouse and human β-globin gene promoters. Naturally occurring single base mutations in the β-globin CACC box result in mild to severe reductions in β-globin gene expression in human thalassemic patients (Orkin et al. 1982, 1984; Kulozik et al. 1991). Similarly, constructs bearing these base substitutions show decreased *trans*-activation of linked reporter genes (Feng et al. 1994). EKLF null mice express endogenous embryonic β-like globins and appear normal during embryonic development (Nuez et al. 1995; Perkins et al. 1995). As the site of erythropoiesis changes to the fetal liver (day 11.5 onward), $EKLF^{-/-}$ animals fail to express the mouse adult β-globin genes and die in utero around day 15. $EKLF^{+/-}$ mice appear normal in terms of β-major and β-minor gene expression (Nuez et al. 1995, Perkins et al. 1995). Because both genes contain CACC boxes in their promoters and are dependent on EKLF for expression (Nuez et al. 1995; Perkins et al. 1995), they are not informative in studies on gene competition (Shehee et al. 1993; T. Trimborn and P. Fraser, unpubl.).

Here we demonstrate the role of EKLF in human β-globin gene switching through analyses of transgenic animals homozygous for a single copy of the entire human β-globin locus in an EKLF heterozygous or homozygous null background. The results show that the absence of EKLF leads to a complete lack of β-gene expression with a concomitant increase in γ gene expression in fetal liver-derived erythroid cells, whereas a decrease in EKLF in heterozygous mice leads to a different ratio of γ to β expression during the switching period. These results indicate that EKLF plays a major role in γ- and β-gene competition and suggest that EKLF is important in the LCR/β-gene interaction.

## Results

### *Embryonic erythropoiesis*

Transgenic mouse line 72 (Strouboulis et al. 1992), which carries a single copy of the complete human β-globin locus, was crossed with a targetted EKLF mutant line (Nuez et al. 1995). Offspring were bred to produce three genotypes for analysis: $Hu\beta^{+/+}/EKLF^{+/+}$ (line 72 homozygous/EKLF wild type), $Hu\beta^{+/+}/EKLF^{+/-}$ (line 72 homozygous/EKLF heterozygous knockout), and $Hu\beta^{+/+}/EKLF^{-/-}$ (line 72 homozygous/EKLF homozygous knockout). Phenotypic analysis of embryonic stage mice suggests that development proceeds normally as EKLF null mutant embryos are indistinguishable from heterozygous or wild-type littermates. S1 nuclease protection analysis of day 10 and 11 blood RNA samples showed that human ε- and γ-globin mRNA expression levels are unchanged in EKLF mutant embryos (Fig. 1A and Fig. 2A) as compared with the unaffected mouse α-globin genes (Nuez et al. 1995; Perkins et al. 1995). Primary transcript in situ hybridization analysis of day 10 embryonic blood cells showed no differences in the number and distribution of active transcriptional foci for human ε and γ globin (Wijgerde et al. 1995) among the three different genotypes (not shown). These results indicate that embryonic expression of the human ε- and γ-globin genes is not dependent on EKLF.

### *Human globin gene expression in the fetal liver*

Around day 11 post conception (p.c.) the fetal liver becomes the major site of erythropoiesis. Expression of human ε globin is completely silenced in the fetal liver, whereas γ-gene expression persists in competition with the β gene (Strouboulis et al. 1992; Wijgerde et al. 1995). $EKLF^{+/-}$ mice develop normally in contrast to $EKLF^{-/-}$ mice, which appear abnormal in terms of hemoglobinization as early as 12.5 days and severely anemic by day 14.5 (Nuez et al. 1995; Perkins et al. 1995). In contrast, $Hu\beta^{+/+}/EKLF^{-/-}$ mice still have hemoglobinized peripheral blood and a relatively red fetal liver as late as 15.5 days (not shown). This result suggests that human β-like chains are heterotetramerizing with mouse alpha chains to form hemoglobin in the fetal liver cells.

S1 protection assays performed on total fetal liver RNA showed dramatic differences in the steady-state globin mRNA levels (Figs. 1B and 2B). The normal pattern of γ- and β-gene expression (Strouboulous et al. 1992) was disrupted in heterozygous EKLF knockout fetuses. Human β-gene expression was nearly halved in the early fetal liver cells (11–15 days) when compared with EKLF wild-type fetuses. The decrease in human β-gene expression was accompanied by an increase in γ-gene expression over the same period. This decrease in human β expression is interesting in light of the fact that the endogenous β-major and β-minor genes are unaffected in EKLF heterozygotes (Fig. 1B; Nuez et al. 1995; Perkins et al. 1995). In $Hu\beta^{+/+}/EKLF^{-/-}$ fetuses human β-gene expression is undetectable and γ-gene expression increases by as much as three or fivefold compared with $Hu\beta^{+/+}/EKLF^{+/-}$ and $Hu\beta^{+/+}/EKLF^{+/+}$ fetuses, respectively. The null mutants also showed a

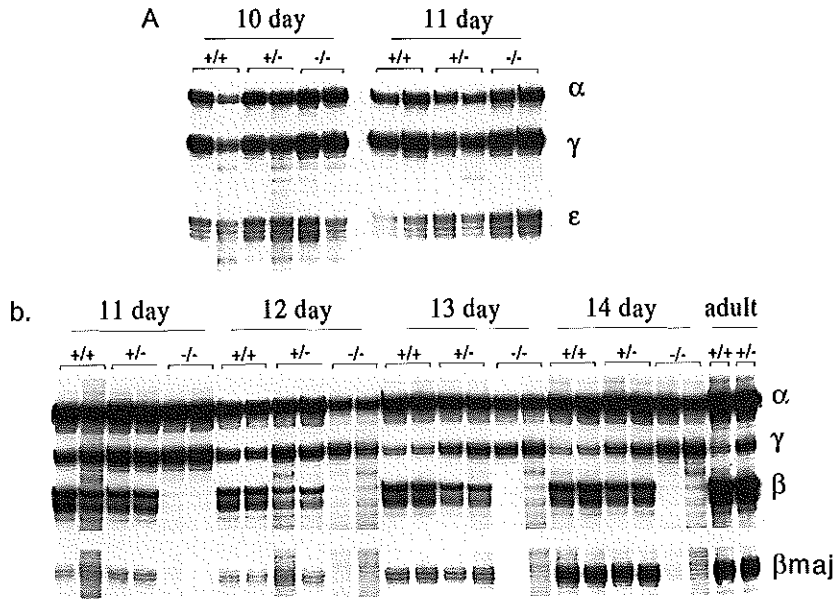



**Figure 1.** S1 nuclease protection assays. Probes for human ε, γ and β globin and mouse α and β-major globin were used as described in Materials and Methods. All mice analyzed were homozygous for a single copy of the entire human β-globin locus [line 72, Strouboulis et al. 1992] in either EKLF$^{+/+}$, EKLF$^{+/-}$, or EKLF$^{-/-}$ background as indicated above each lane. (*A*) Total RNA from 10- and 11-day embryos including yolk sac. (*B*) Eleven- to 14-day fetal livers and adult peripheral blood RNA. Protected products are indicated to the *right* of each panel.

dramatic decrease in the level of mouse β-globin mRNA, as expected [Nuez et al. 1995; Perkins et al. 1995].

There are two potential explanations for the observed increase in γ- and decrease in β-mRNA expression. It is possible that the reduced amount of EKLF in the heterozygous knockout animals abates the transcriptional activity of the β genes by reducing polymerase density and, therefore, rate of transcription of all β genes. However, reduced EKLF levels must also have the opposite effect on the transcription rate of individual γ genes to increase expression. EKLF would then be a purported suppressor of γ-gene transcription, or indirectly affect such a gene product, which we cannot rule out. However, it is known that EKLF is present in embryonic red cells where its presence does not silence γ genes. The other possibility, which fits with our previous results on gene competition, would suggest that competition from the γ gene for the LCR is more successful in an EKLF$^{+/-}$ background as a result of decreased competence of the β gene. In this situation the frequency of LCR–γ gene interactions would be increased because of a decrease in the duration [stability] or frequency of LCR–β gene interactions. This would result in a decrease in the absolute number of transcriptionally active β genes within the population and a reciprocal increase in the number of transcriptionally active γ genes. This is supported by the return to near normal levels of human β-globin expression in the adult blood of EKLF$^{+/-}$ mice when the γ genes are silenced. To differentiate between these two possibilities we quantitated the number of transcriptionally active γ- and β-globin genes in the fetal liver.

### *Primary transcript in situ hybridization*

The analysis of steady-state mRNA levels for human γ and β globin during fetal liver erythropoiesis demonstrated concomitant increases in γ-gene expression with decreases in β-gene expression, confirming the reciprocal relationship in the competition between the genes. Our previous results suggested that if a globin gene was transcriptionally active, it was transcribed at the full rate because changes in mRNA levels could be correlated with changes in the number of transcriptionally active genes [Wijgerde et al. 1995; Milot et al. 1996]. Fifty percent expression of a gene at the mRNA level did not suggest that each gene was transcribed at half the normal rate, but rather that half the genes were transcribed at the normal rate [Wijgerde et al. 1995; Milot et al. 1996]. In other words, each gene was transcribed for half the time. To further investigate gene competition and γ-gene

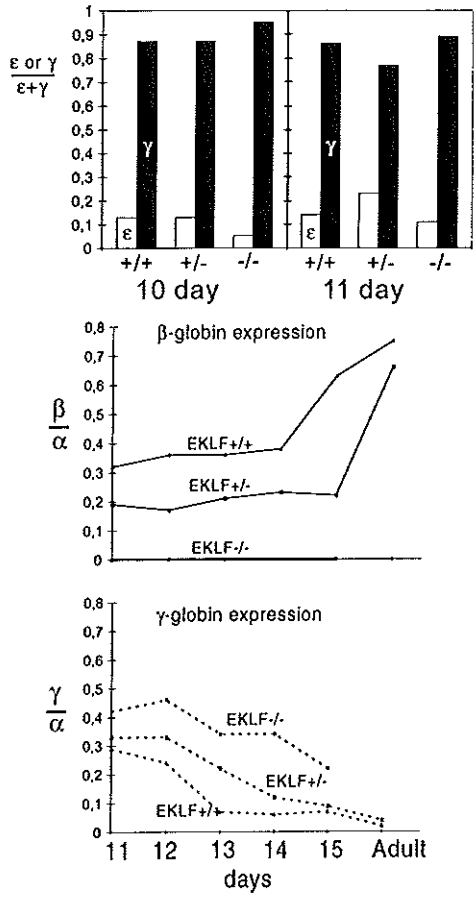



**Figure 2.** Expression pattern of the human globin transgenes in EKLF null background. Expression levels were quantitated using PhosphorImager analysis of S1 protection assays in Fig. 1, and plotted against the developmental age of the animals. (*A*) Relative expression levels of ε- and γ-globin mRNA in day 10 and day 11 whole embryos in EKLF$^{+/+}$, EKLF$^{+/-}$, or EKLF$^{-/-}$ background. (*B*) Expression of human β-globin mRNA during fetal stages of erythropoiesis and in adult peripheral blood as a ratio of mouse α-globin mRNA. (*C*) Expression of human γ-globin mRNA during fetal stages of erythropoiesis and in adult peripheral blood as a ratio of mouse α-globin mRNA.

silencing we analyzed transcription of individual γ- and β-globin genes in single cells using double label primary transcript in situ hybridization (Wijgerde et al. 1995) in 12.5-day fetal liver cells (Fig. 3A and B). The results show that the percentage of transcription foci for the β gene decreases from 80% in the wild type to 52% in EKLF$^{+/-}$ and 0% in EKLF$^{-/-}$ mice (Fig. 4 and Table 1). γ-gene transcription foci increase from 20% in wild type to 48% in EKLF$^{+/-}$ fetuses. Because no human β-globin gene foci were observed in the EKLF$^{-/-}$ fetuses we also compared γ-gene foci with mouse α-globin gene foci in 13.5-day wild type, EKLF$^{+/-}$, and EKLF$^{-/-}$ fetuses (Table 1 and Fig. 3C and D). The results show that in EKLF$^{-/-}$ mice only 41% of the γ genes in the red cells were transcriptionally active, suggesting that many γ genes already may have been irreversibly silenced (Dillon and Grosveld 1991, see Discussion). Correlation of the relative percent of γ and β mRNA in the fetal liver population (S1 analysis) and the number of transcriptionally active γ and β genes (in situ data) indicates that changes in the expression levels of the individual genes are wholly accounted for by changes in the number of transcriptionally active genes (Fig. 4 and Table 1) and not attributable to changes in the rate of transcription. These results confirm that the level of expression of a globin gene is determined by the frequency and duration of its transcription and not by changes in the rate of transcription or polymerase density along a single template. Hence we conclude a gene is either transcribed fully or is silent.

*DNaseI hypersensitive site analysis*

S1 nuclease protection assays combined with primary transcript in situ hybridization have revealed that the

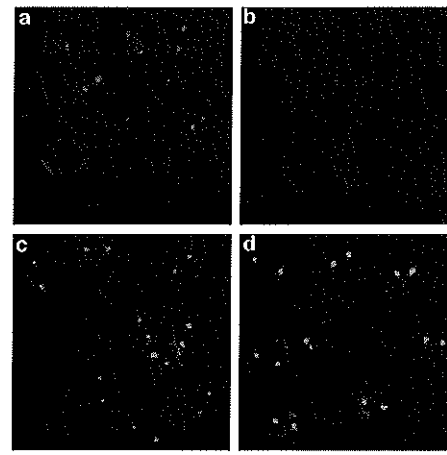



**Figure 3.** Primary transcript in situ hybridization. EKLF wild-type and knockout 13.5-day mouse fetal liver cells homozygous for the human β-globin locus were hybridized with gene-specific intron probes for (*A,B*) human γ (red) and β globin (green) or (*C,D*) mouse α globin (green) and human γ globin (red). *A* and *B* were created with a CCD camera. *C* and *D* were created with a laser scanning confocal microscope. Genotypes (*A*) Huβ$^{+/+}$ EKLF$^{+/+}$; (*B*) Huβ$^{+/+}$ EKLF$^{-/-}$; (*C*) Huβ$^{+/+}$ EKLF$^{+/+}$; (*D*) Huβ$^{+/+}$ EKLF$^{-/-}$.

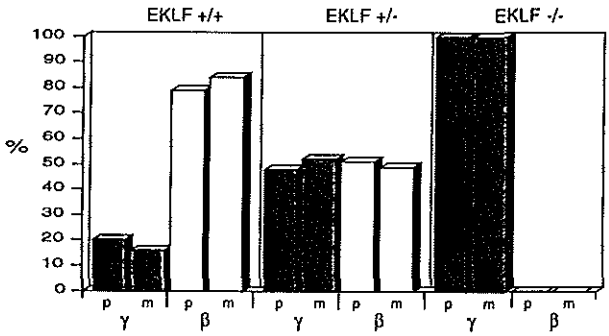



**Figure 4.** Correlation between transcriptional activity and steady-state mRNA levels. Quantities of steady-state mRNA (γ or β) at 13.5 days postconception from Figs. 1 and 2 are represented as a percentage of total human RNA (γ or β/γ + β). The percentages of transcriptionally active γ and β genes were calculated from the primary transcript in situ data (Table 1; 13.5 days) by dividing the number of transcriptionally active γ or β genes by the total number of transcriptionally active γ and β genes (γ or β/γ + β). Loci with double signals were equally divided between γ and β categories. (p) Primary transcript foci; (m) mRNA levels. Values for EKLF$^{+/+}$, EKLF$^{+/-}$, and EKLF$^{-/-}$ are shown as indicated.

decrease in the steady-state level of human β-globin mRNA is a result of a decrease in the number of actively transcribing β genes in the EKLF$^{+/-}$ and EKLF$^{-/-}$ mice resulting in a concomitant increase in the number of transcriptionally active human γ-globin genes. Globin gene transcription normally correlates with the presence of DNase I hypersensitive sites at the promoter and in the LCR (Tuan et al. 1985; Forrester et al. 1987; Grosveld et al. 1987). To determine the effect of reduced EKLF levels on chromatin structure in the human and mouse β-globin loci, we analyzed DNase I hypersensitivity of promoter and LCR HS regions in isolated nuclei from wild-type and EKLF$^{-/-}$ 13.5-day fetal liver cells. As expected, no difference was seen in the hypersensitivity of the unaffected mouse α-globin gene promoters (Fig. 5). In the absence of EKLF most of the globin locus appears normal in terms of DNase I hypersensitivity (Table 2); however, the adult β-globin promoter is resistant to DNase I digestion (Fig. 5). The mouse β-major globin promoter is also resistant to DNase I and the mouse and human LCR HS3 appear to be reduced to half of normal sensitivity (Fig. 5 and Table 2). HS:1, 2, and 4 showed no difference in sensitivity between wild-type and knockout mice (Table 2). These results suggest that EKLF primarily affects chromatin structure at the adult β-globin promoter and to a lesser extent HS3 of the LCR in both

**Table 1.** *Quantitation of in situ hybridizations*

| | γ | β | γ–β | γ/α |
|---|---|---|---|---|
| EKLF +/+ | 13 | 72 | 15 | 18 |
| EKLF +/− | 37 | 41 | 22 | 33 |
| EKLF −/− | 100 | 0 | 0 | 41 |

The percentages of loci in 13.5-day fetal liver cells having single gene primary trancript signals for γ or β, and those having double signals γ–β are shown for the three genotypes indicated. Greater than 170 loci were counted for each genotype. γ/α represents the number of transcriptionally active γ genes as a percentage of transcriptionally active mouse α-globin genes. Greater than 500 α-globin loci were counted for each genotype.

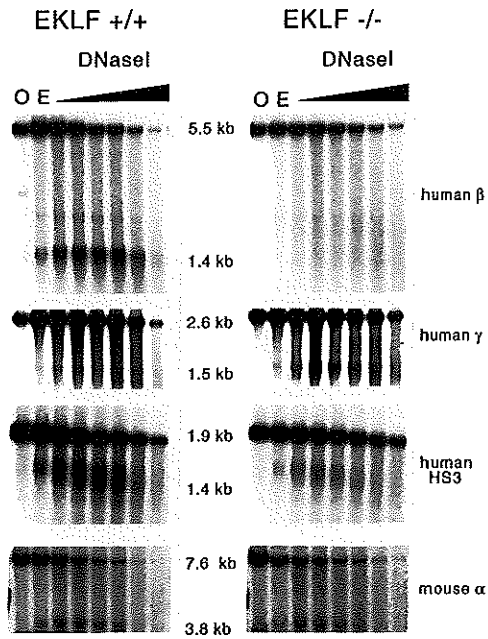



**Figure 5.** DNase I hypersensitive site analysis of 13.5-day fetal liver cells. Shown are DNase I hypersensitivity analyses in isolated nuclei from fetal livers of mice homozygous for the human β-globin locus in EKLF$^{+/+}$ and EKLF$^{-/-}$ backgrounds. The probes used are described in Materials and Methods and the sizes of parent and hypersensitve bands are indicated. Shown from *top* to *bottom*: Human β-globin promoter (*top* panel, *Eco*RI digest), human γ-globin promoter (*Eco*RI digest), human LCR HS 3 (*Hin*dIII digest), and endogenous mouse α-globin promoter (*bottom* panel, *Sac*I digest). Lanes labeled with O and E indicate samples incubated on ice and 37°C, respectively, without DNase I. The triangled bar above the lanes indicates increasing amounts of DNase I.

Table 2. *DNaseI hypersensitive site analysis performed on line72 EKLF$^{+/+}$ and line72 EKLF$^{-/-}$ 13.5-day fetal liver cell isolated nuclei*

| | DNaseI hypersensitive sites | |
|---|---|---|
| | EKLF$^{+/+}$ | EKLF$^{-/-}$ |
| Human β promoter | + | − |
| Human γ promoter | − | + |
| Human LCR HS1 | + | + |
| Human LCR HS2 | + | + |
| Human LCR HS3 | + | +/− |
| Human LCR HS4 | + | + |
| Murine βmaj promoter | + | − |
| Murine LCR HS3 | + | +/− |
| Murine α promoter | + | + |

For detailed description of the procedure and probes used, see Materials and Methods.

the endogenous mouse and human β-globin loci. Of particular interest is the observation that DNase I hypersensitivity of the human γ-globin gene promoter is increased in EKLF$^{-/-}$ mice compared with wild-type mice as the number of actively transcribed γ-genes increases (Fig. 5).

## Discussion

### *Modulation of gene expression*

We have shown that the level of human β-globin mRNA in the EKLF$^{+/-}$ and EKLF$^{-/-}$ mice decreases as a direct consequence of decreases in the number of actively transcribing β-globin genes. The results show reciprocal concomitant increases in human γ-globin gene expression resulting solely from increases in the number of transcriptionally active human γ-globin genes rather than increases in the rate of γ-gene transcription on individual genes. These data indicate that gene transcription is an all-or-nothing event and that changes in the level of gene expression reflect changes in the amount of time that a gene is transcribed rather than changes in the rate of transcription of individual genes. These observations support our previous work, which showed that as development proceeds, modulation of the expression level of γ versus β mRNA is determined by the frequency and duration of transcription of the individual genes and not by changes in the rate of transcription from individual promoters (Wijgerde et al. 1995). The genes appeared to be either fully transcribed or off. This concept was independently reinforced by data that showed that novel position effects that cause low-level expression in all cells of a tissue are the result of decreasing the amount of time that a gene is transcriptionally active and not the result of modulation of transcription rate or polymerase density on an individual gene (Milot et al. 1996).

Other groups have proposed a binary mechanism of enhancer action. Robertson et al. (1995) described the silencing of expression of a construct containing an *Escherichia coli lacZ* gene under control of the human α-globin HS-40 element in transgenic mice. It had been shown previously that HS-40 could not maintain high-level copy number-dependent expression of the α-globin gene throughout development as expression levels declined as development proceeded (Sharpe et al. 1992, 1993). Histochemical staining for β-gal showed that the lower expression levels seen late in development correlated with decreasing numbers of *lacZ*-expressing cells, and suggested that a decrease in the rate of transcription from individual promoters did not play a role. A report by Walters et al. (1996) described silencing of an HS:2 β-geo construct in cell culture after recombinase mediated removal of HS:2. They found that after culturing stably transfected clones for several weeks, greater numbers of cells, though variable from line to line, had silenced the β-geo gene when compared with the original HS:2-containing clone. It was suggested that the level of β-gal expression per cell did not change dramatically in those cells that continued to express the β-geo gene, although twofold variation was common. They concluded that enhancers act to prevent silencing of genes and have little if anything to do with the rate of transcription of a gene. In addition, it was concluded that in terms of transcription, a gene was either stably on or off in a particular cell, although in neither report was transcription measured nor was the activation/silencing status of a cell heritable as in position effect variegation (PEV). Both reports show that levels of steady-state mRNA extracted from a population of cells do not always reflect changes in all of the cells. We have observed a similar phenomenon in cases of genuine PEV in which a stable, clonal subpopulation of cells expresses or silences a transgene as a result of juxtaposition to centromeric heterochromatin leading to reductions in mRNA levels in a population of cells (Milot et al. 1996). However, these types of experiments do not allow conclusions to be made about modulation of gene expression in individual expressing cells. We have shown that expression levels of individual genes are modulated within a single cell by altering the periods of active transcription of a gene and not the rate of transcription (Wijgerde et al. 1995; Milot et al. 1996; this report). This result on the dynamics of transcription is fundamentally different from the work on gene silencing mentioned above. Gene transcription appears to be an all-or-nothing event, and modulation of gene expression within a single cell is accomplished through altering the transcriptional periods.

### *Shifting the balance of globin gene competition*

The analyses of human embryonic β-like globin gene expression and transcription in EKLF null mice demonstrate that the ε- and γ-gene promoters are not dependent on EKLF. Similar results were obtained when the mouse embryonic globin genes were analyzed (Nuez et al. 1995; Perkins et al. 1995). Human globin gene expression in the fetal liver of transgenic mice involves a prolonged period of gene competition between the γ- and β-globin genes from day 11.5 to approximately day 16.5 (Strouboulis et al. 1992; Peterson and Stamatoyannopo-

ulos 1993; Wijgerde et al. 1995). During this period it is thought that a gradual change in the transcription factor environment of individual cells leads to changes in the affinity and/or stability of interactions between the LCR and the individual genes, causing a shift in the balance of transcription toward the adult β gene. We have shown previously that expression of multiple globin genes from a single locus involves alternating transcription of individual genes (Wijgerde et al. 1995). We proposed a mechanism whereby the LCR forms an exclusive semistable complex with an individual gene to initiate transcription but is able to flip-flop between genes. The implication of the proposed mechanism is that formation of a stable complex between the LCR and a gene allows loading or initiation of RNA polymerases (and reinitiation) at a fixed rate. The frequency and stability of the complex would therefore determine the level of transcription and disruption of the interaction would prevent further loading of polymerases. Support for the idea that the LCR acts as a functional unit has come from studies that show that all HS of the LCR are required for position-independent and copy number-dependent expression (Milot et al. 1996). Indeed, deletion of individual HS, which one might assume would destabilize the complex, leads to novel postion effects in which transgene expression levels were lower as a result of reduced transcription time (Milot et al. 1996).

Here we show that in EKLF$^{+/-}$ mice the ratio of γ- to β-gene transcription is shifted toward the γ genes when compared with wild-type mice. This shift in the ratio represents an actual increase in the absolute number of actively transcribed γ genes and a decrease in the number of transcribed β genes. Therefore the decreased level of EKLF in heterozygous knockouts leads to a reduction in the amount of time that the LCR spends complexed with the β gene. Because the γ-gene promoter is not directly affected by EKLF we propose that increased γ-gene transcription is the result of increased availability of the LCR as a result of reduction in the time it is occupied by the β gene. Hence LCR–γ gene interactions, though unchanged in stability, are able to form more frequently.

Martin et al. (1996) have described a different model of globin gene regulation. They propose that the LCR acts in concert with elements in the locus to initiate and maintain an active chromatin structure. No contact or interaction is required between the LCR and the genes in the locus for transcriptional activation. Instead, individual genes within the locus have differing probabilities of transcriptional activation depending on gene-local *cis*-elements and the developmentally regulated *trans*-acting factors that bind them. Multiple genes may be transcribed simultaneously from the same locus but 5′ genes may interfere with the transcription of 3′ genes by either lowering the probability or decreasing the rate of transcription. Silencing of the more 5′ genes during development allows activation of the downstream genes. However, the data presented here directly contradicts this model as we have shown that decreases in transcription of a 3′ gene (the β gene) lead to reciprocal increases in a 5′ gene (the γ genes). These results are exactly what one would predict in a case of genuine gene competition, and are consistent with the observations that the switch from γ to β expression is competitive (Hanscombe et al. 1991; Peterson and Stamatoyannopoulos 1993; Dillon et al. 1995; N. Dillon and F. Grosveld, in prep.) and dynamic (Wijgerde et al. 1995; this report).

### *γ-gene silencing*

In the homozygous knockout animals β-gene transcription is not detectable and the percentage of loci actively transcribing the γ-genes increases to 41% of transcriptionally active mouse α-globin loci (Fig. 3 and Table 1). Of interest is the fact that not all γ genes are transcribed in the absence of β-gene transcription. It is known that γ-gene transcription is silenced autonomously during fetal liver development based on expression analysis of a γ-gene linked to the LCR (Dillon and Grosveld 1991). Our observations would predict that the decreasing level of γ-gene expression seen in those experiments is the result of decreasing numbers of transcriptionally active γ-genes (i.e., increasing numbers of silenced loci). One possible explanation is that in the EKLF $^{-/-}$ mice there is a distinct subset of fetal liver cells in which the γ-genes are still capable of transcription (i.e., not silenced). Our results show that the proportion of cells with both loci transcribing the γ genes increases significantly in the double knockout mice compared with wild type and heterozygotes at the expense of cells with only one locus active in γ-gene transcription. However, even in the homozygous knockout the majority of cells that are transcribing γ have only one active locus. It is possible that the other locus appears inactive because the LCR still spends a significant, though reduced amount of time complexed with the β-gene but does not lead to productivity. An alternative explanation is that the γ genes on the other locus are reacting independently to the same factor environment (as observed previously; Wijgerde et al. 1995) and have been silenced thereby making them unavailable for LCR interaction and transcription. Comparison with the results of Dillon et al. (1991) in which γ silencing was measured in the absence of competition from the β-gene suggests that we may be observing a combination of the possibilities described above.

### *Chromatin structure*

The results of DNase I hypersensitive site analysis suggest that the chromatin structure of the human globin LCR remains largely intact in EKLF double knockout mice. HS 1, -2, and -4 of the LCR appear unchanged in the absence of EKLF. The decrease in sensitivity of the promoter of the adult β genes (human and mouse) that is coincident with transcriptional deficiency is quite severe, indicating that EKLF is essential for the proper chromatin structure of the promoter. Hypersensitivity of the γ-gene promoter increases in the EKLF double knockout mice as transcription increases. Our model predicts that this increase is the result of an increase in

the number of γ genes interacting with the LCR. It is not possible to determine with certainty whether hypersensitivity is the result of transcription of the gene or interaction with the LCR, as the two are normally linked. If the LCR is complexed with the β gene in the homozygous knockouts as suggested above, one would then conclude that hypersensitivity is more closely linked to transcription than LCR interaction. EKLF also contributes, to a lesser extent, to the hypersensitivity of HS 3 in the mouse and human LCRs, suggesting that it may directly interact with sequences in HS 3.

## Materials and methods

### *Transgenic mice*

Transgenic mice containing a single integrated copy of the human β-globin locus, referred to as Line 72 (Strouboulis et al. 1992), were crossed with EKLF knockout mice (Nuez et al. 1995) to create compound homozygous human β-globin Huβ$^{+/+}$/EKLF$^{+/-}$ and Huβ$^{+/+}$/EKLF$^{-/-}$ embryos. Embryos were dissected out at various time points and genotyped by Southern blot and PhosporImage analysis.

### *Preparation of RNA and S1 nuclease protection assay*

RNA was extracted from frozen embryos (including yolk sac), fetal livers, and adult blood (animals >8 weeks old) and analyzed by S1 nuclease protection assays as described previously (Fraser et al. 1990). Quantitation of signals was performed using a PhosphorImager (Molecular Dynamics). S1 probes (Lindenbaum and Grosveld 1990; Milot et al. 1996) were end-labeled with T4 polynucleotide kinase and equimolar amounts of each end-labeled probe were used per reaction. Amounts of RNA analyzed per protection assay: day 10 and 11 embryos, 6 μg; day 11 fetal livers, 3 μg; day 12 fetal livers, 2 μg; day 13, 14, 15, and 16 fetal livers, 1 μg; adult blood, 1 μg.

### *Primary transcript in situ hybridization*

Embryonic blood (10.5-day) and 12.5- and 13.5-day fetal livers were disrupted in PBS. Cells were spotted and immobilized onto poly-L-lysine-coated slides and subsequently fixed in 4% formaldehyde, 5% acetic acid in saline for 20 min at room temperature. Slides were further processed and used for in situ hybridization analysis and antibody detection as described previously (Wijgerde et al. 1995). Probes were labeled with digoxygenin, biotin, or dinitrophenol. Transcription signals were quantitated by counting cells (see Table 1) using an epifluorescence microscope.

### *Isolation of nuclei and DNase I hypersensitive site analysis*

Nuclei were isolated from 13.5-day frozen fetal livers as described by Forrester et al. (1990). For each genotype (EKLF$^{+/+}$, EKLF$^{-/-}$) 10 livers were disrupted with 10 strokes of a type B pestle. Aliquots (100-μl) were digested with increasing amounts of DNase I for 3 min at 37°C. Two aliquots, one incubated on ice and one at 37°C, were devoid of any DNase I and served as controls for the zero time point and endogenous nucleases, respectively. Reactions were treated with 250 μg/ml proteinase K in 0.3 M NaCl, 0.5% SDS, 5 mM EDTA, and 10 mM Tris at pH 8 at 55°C for 16 hr, phenol/chloroform extracted and ethanol precipitated. Pellets were dissolved in 80 μl of distilled water. Twenty microliters of DNA was digested with appropriate restriction enzyme and analyzed via Southern blot. The following restriction enzymes and probes were used: γ promoter, *Eco*RI digest, *Bam*HI–*Eco*RI γIVSII fragment; β promoter, *Eco*RI digest, *Bam*HI–*Eco*RI βIVSII fragment; human HS1, *Pst*I, 520-bp *Hin*cII fragment; human HS2, *Pst*I, 1.5-kb *Kpn*I–*Bgl*II; human HS3, *Hin*dIII, 1.9-kb *Hin*dIII; human HS4, *Hin*dIII digest, 600-bp *Bam*HI–*Sac*I; β-maj promoter, *Eco*RI, 1-kb *Eco*RI–*Hin*dIII; mouse HS3, *Hin*dIII digest, 1.9-kb *Hin*dIII; mouse α1 and α2, *Sac*I digest, 600-bp *Bam*HI–*Sac*I. The relative hypersensitivity between knockout and wild-type lines was quantitated via PhosphorImage analysis. The ratio of hypersensitive bands over parental band was plotted and the slopes of the resulting lines were compared.


## Acknowledgments

We are grateful to Eric Milot for helpful discussions and comments, to Lien Braam and Corinne Stark for animal care, and to Adriaan Houtsmuller for confocal assistance. B.N. was supported by a Human Frontiers Fellowship. This work was supported in part by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands) and the Howard Hughes Foundation.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.


## References


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

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

Dillon, N., J. Strouboulis, and F. Grosveld. 1995. The reglation of human β-globin gene expression: Polarity of transcriptional competition in the human β-globin locus. In *Molecular biology of hemoglobin switching*, Proceedings of the 9th Conference on Hemoglobin Switching held at Orcas Island, Washington, USA (ed. G. Stamatoyannopoulous), pp. 23–28. Intercept Limited, Andover, Hants, UK.

Donze, D., T.M. Townes, and J.J. Bieker. 1995. Role of erythroid Kruppel like factor in human γ- to β-globin gene switching. *J. Biol. Chem.* **270:** 1955–1959.

Feng, W.C., C.M. Southwood, and J.J. Bieker. 1994. Analysis of β-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J. Biol. Chem.* **269:** 1493–1500.

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

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

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

Fraser, P., S. Pruzina, M. Antoniou, and F. Grosveld. 1993. Each hypersensitive site of the human beta-globin locus control

region confers a different developmental pattern of expression to the globin genes. *Genes & Dev.* **7:** 106–113.

Grosveld, F., G. Blom van Assendelft, D.R. Greeves, and G. Kollias. 1987. Position independent, high level expression of the human β-globin gene in transgeneic mice. *Cell* **51:** 975–985.

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

Kulozik, A.E., A. Bellan-Koch, S. Bail, E. Kohne, and E. Kleihauer. 1991. A deletion/inversion rearrangement of the beta-globin gene cluster in a Turkish family with delta beta zero-thalassemia intermedia. *Blood* **77:** 2054–2058.

Lindenbaum, M.H. and Grosveld, F. 1990. An in vitro globin gene switching model based on differentiated embryonic stem cells. *Genes & Dev.* **4:** 2075–2085.

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

Miller, I.J. and J.J. Bieker. 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.

Milot, E., J. Strouboulis, T. Trimborn, M. Wijgerde, E. de Boer, A. Langeveld, K. Tan-Un, W. Vergeer, N. Yannoutsos, F. Grosveld, and P. Fraser. 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87:** 1–10.

Nuez, B., D. Michalovich, A. Bygrave, R. Ploemacher, and F. Grosveld. 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* **375:** 316–318.

Orkin, S.H., H.H. Kazazian, Jr., S.E. Antonarakis, S.C. Goff, C.D. Boehm, J.P. Sexton, P.G. Waber, and P.J. Giardina. 1982. Linkage of beta-thalassaemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human beta-globin gene cluster. *Nature* **296:** 627–631.

Orkin, S.H., S.E. Antonarakis, H.H. Kazazian, Jr. 1984. Base substitution at position -88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. *J. Biol. Chem.* **259:** 8679–8681.

Perkins, A.C., A.H. Sharpe, and S.H. Orkin. 1995. Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* **375:** 318–322.

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

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

Robertson, G., D. Garrick, W. Wu, M. Kearns, D. Martin, and E. Whitelaw. 1995. Position-dependent variegation of globin transgene expression in mice. *Proc. Natl. Acad. Sci.* **92:** 5371–5375.

Sharpe, J.A., P.S. Chan-Thomas, J. Lida, H. Ayyub, W.G. Wood, and D.R. Higgs. 1992. Analysis of of the human α-globin upstream regulatory element (HS-40) in transgenic mice. *EMBO J.* **11:** 4565–4570.

Sharpe, J.A., D.J. Wells, E. Whitelaw, P. Vyas, D.R. Higgs, and W.G. Wood. 1993. Analysis of the human α-globin gene cluster in transgenic mice. *Proc. Natl. Acad. Sci.* **90:** 11262–11266.

Shehee, R., P. Oliver, and O. Smithies. 1993. Lethal thalassemia after insertional disruption of the mouse major adult beta-globin gene. *Proc. Natl. Acad. Sci.* **90:** 3177–3181.

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

Tuan, D., W. Salomon, Q. Li, and I. London. 1985. The "β-like globin" gene domain in human erythroid cells. *Proc. Natl. Acad. Sci.* **82:** 6384–6388.

Walters, M.C., W. Magis, S. Fiering, J. Eidemiller, D. Scalzo, M. Groudine, and D.I.K. Martin. 1996. Transcriptional enhancers act in cis to suppress position-effect variegation. *Genes & Dev.* **10:** 186–195.

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

# Chapter 4

Erythroid Krüppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5′HS3 of the β-globin locus control region

*Tewari, R., Gillemans, N., Wijgerde, M., von Lindern, M., Grosveld, F. and Philipsen, S.*
*EMBO journal 17, 2334-2341, (1998)*

# Erythroid Krüppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5′HS3 of the β-globin locus control region

**Rita Tewari[1], Nynke Gillemans[2], Mark Wijgerde[2], Beatriz Nuez[3], Marieke von Lindern[4], Frank Grosveld[2,5] and Sjaak Philipsen[2]**

Erasmus University Rotterdam, MGC, [2]Department of Cell Biology and [4]Department of Haematology, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

[1]Present address: Sars International Centre for Molecular Marine Biology, Bergen High Technology Centre, Thormøhlensgt. 55, N-5020 Bergen, Norway
[3]Present address: Chandika DP Towers, Apartment 7, 8 College Lane, Nungambakkam, 600 006 Tamil Nadu, Channai, India

[5]Corresponding author

**Disruption of the gene for transcription factor EKLF (erythroid Krüppel-like factor) results in fatal anaemia caused by severely reduced expression of the adult β-globin gene, while other erythroid-specific genes, including the embryonic ε- and fetal γ-globin genes, are expressed normally. Thus, EKLF is thought to be a stage-specific factor acting through the CACC box in the β-gene promoter, even though it is already present in embryonic red cells. Here, we show that a β-globin gene linked directly to the locus control region (LCR) is expressed at embryonic stages, and that this is only modestly reduced in EKLF$^{-/-}$ embryos. Thus, embryonic β-globin expression is not intrinsically dependent on EKLF. To investigate whether EKLF functions in the locus control region, we analysed the expression of LCR-driven *lacZ* reporters. This shows that EKLF is not required for reporter activation by the complete LCR. However, embryonic expression of reporters driven by 5′HS3 of the LCR requires EKLF. This suggests that EKLF interacts directly with the CACC motifs in 5′HS3 and demonstrates that EKLF is also a transcriptional activator in embryonic erythropoiesis. Finally, we show that overexpression of EKLF results in an earlier switch from γ- to β-globin expression. Adult mice with the EKLF transgene have reduced platelet counts, suggesting that EKLF levels affect the balance between the megakaryocytic and erythroid lineages. Interestingly, the EKLF transgene rescues the lethal phenotype of EKLF null mice, setting the stage for future studies aimed at the analysis of the EKLF protein and its role in β-globin gene activation.**
*Keywords:* EKLF/β-globin/locus control region/primitive erythropoiesis

## Introduction

The human β-globin gene cluster contains five functional genes 5′-ε (embryonic)-Gγ-Aγ (fetal)-δ-β (adult)-3′. Expression of these genes is completely dependent on the presence of the locus control region (LCR) upstream of the ε-globin gene. The LCR contains five DNaseI hypersensitive sites, termed 5′HS1–5, distributed over 20 kb, and a fully functional 6.5 kb version of the LCR containing 5′HS1–4 (μLCR) has been described (Talbot *et al.*, 1989). The functional activity of the 5′HSs resides in 200–300 bp 'core' fragments coinciding with the hypersensitive areas (reviewed in Grosveld *et al.*, 1993).

Erythroid Krüppel-like factor (EKLF) is an erythroid-specific transcription factor containing three archetypal zinc fingers (Miller and Bieker, 1993). It binds to the DNA sequence CCA CAC CCT, which is found in the promoter of the adult β-globin gene, and related sequences (Feng *et al.*, 1994). Naturally occurring point mutations in this promoter element, also known as the CACC box, result in reduced expression of the β-globin gene linked *in cis* (reviewed in Thein, 1993), suggesting that EKLF is an important activator of β-globin gene expression. This notion was confirmed by the phenotype of EKLF knock-out mice (Nuez *et al.*, 1995; Perkins *et al.*, 1995). These mice die at around day 14–15 of gestation (E14–15) due to anaemia caused by failure to express the β-globin gene.

The role of EKLF in the expression of the human β-globin gene cluster has been analysed by crossing mice carrying the complete human locus into the EKLF null background (Perkins *et al.*, 1996; Wijgerde *et al.*, 1996). This showed that the embryonic ε- and the fetal γ-globin genes are expressed at normal levels in EKLF$^{-/-}$ fetuses, but expression of the adult β-globin gene is barely detectable. Thus, these data suggest that EKLF exerts its effect on β-globin expression through the CACC box in the promoter of this gene. Hypersensitivity analysis showed that the 5′HS3 region of the LCR was also affected. It contains CACC box motifs, suggesting that the loss of hypersensitivity is a direct result of the absence of EKLF. However, the CACC box is also present in the promoters/enhancers of many other erythroid-specific genes including the α-globins (Philipsen *et al.*, 1990) which are not affected. Thus, the loss of hypersensitivity could be an indirect effect caused by changes in the interactions of the LCR with the β-globin gene (Wijgerde *et al.*, 1996).

In the developing mouse embryo, the first erythroid cells are derived from the yolk sac and start to appear at E7.5. These cells remain nucleated and are referred to as primitive cells. The first enucleated erythrocytes are formed in the fetal liver at around E11.5; both fetal and adult erythrocytes are referred to as definitive cells. There is good evidence that primitive and definitive cells arise from two independent lineages in the embryo (reviewed in Dzierzak and Medvinsky, 1995). Interestingly, the EKLF protein is already present in embryonic erythroid cells (Southwood *et al.*, 1996), but as yet, no defect in primitive erythropoiesis has been reported for EKLF$^{-/-}$ mice (Nuez

*et al.*, 1995; Perkins *et al.*, 1995). Thus, it would appear that EKLF is not active in primitive cells.

In this paper, we have investigated the activity of EKLF in the developing erythroid system and its potential role in the LCR. First, we demonstrate that a β-globin gene linked directly to the μLCR is expressed at the embryonic stage and that embryonic expression is only moderately affected in EKLF$^{-/-}$ embryos. Thus, expression of the β-globin gene is not intrinsically dependent on EKLF in primitive cells. Secondly, we investigated the link between EKLF and LCR activity in the absence of β-globin gene sequences through the use of μLCR-*lacZ* mice (Tewari *et al.*, 1996). We show that embryonic expression of the μLCR-*lacZ* construct is retained in the EKLF null background. In contrast, the expression of *lacZ* constructs with only 5'HS3 or the core fragment of 5'HS3 is dependent on EKLF in primitive cells, demonstrating that EKLF is also active in primitive cells and required for the activity of 5'HS3 of the LCR. Next, we investigated the effect of increased EKLF levels on γ- to β-globin switching by expressing EKLF cDNA under the control of the β-globin LCR and promoter. The results show that increased EKLF levels change the balance from γ- to β-globin gene expression in definitive cells, suggesting that the amount of EKLF influences the rate of the switching process. We further used the EKLF transgene to rescue the EKLF null mutation. These rescue mice had almost normal haematological parameters, with slightly reduced haemoglobin levels due to a reduction in the number of erythrocytes. Interestingly, wild-type mice with the EKLF transgene have reduced numbers of circulating platelets, suggesting that EKLF may have a role in determining the balance between the megakaryocytic and erythroid lineages.

## Results

### *Expression of the β-globin gene in primitive erythropoiesis*

A β-globin gene linked directly to the LCR is expressed prematurely in embryonic erythroid cells (Hanscombe *et al.*, 1991; Dillon *et al.*, 1997). The EKLF protein is known to be present in these cells (Southwood *et al.*, 1996) but no function has been assigned to EKLF in primitive erythropoiesis (Nuñez *et al.*, 1995; Perkins *et al.*, 1995); a Northern blot showing the absence of EKLF-encoding mRNA in E10.5 –/– yolk sac RNA is shown in Figure 1A. Thus, we wondered whether embryonic expression of the β-globin gene is also dependent on EKLF. To test this directly, we used a transgenic line with a single copy of the μLCR-β-globin construct (Figure 1B). In E13.5 fetal liver, this human β-globin transgene is expressed at the same level as the endogenous β-major gene (line μD14 in Ellis *et al.*, 1996). First, we determined transgene expression in primitive cells. We isolated RNA from E10.5 embryos and quantitated the expression level of the human β-globin gene relative to that of the endogenous mouse α-globin gene, which is expressed in both primitive and definitive cells and is not affected by the EKLF knock-out. The results are shown in Figure 1C. As could be predicted, the β-globin gene is expressed prematurely due to the absence of the ε- and γ-globin genes in the construct (Hanscombe *et al.*, 1991; Dillon *et al.*, 1997). We then analysed expression of the transgene in EKLF$^{-/-}$ embryos. Interestingly, there is only a modest (2- to 3-fold), reduction in the level of transgene-derived mRNA in E10.5 EKLF$^{-/-}$ embryos, while expression of the transgene is more drastically affected in E13.5 EKLF$^{-/-}$ fetal liver (>6-fold reduction). This reduction is smaller than reported previously for adult β-globin expression (Perkins *et al.*, 1996; Wijgerde *et al.*, 1996) which could be the result of the presence of human β-globin mRNA in circulating primitive cells. We therefore performed primary transcript *in situ* hybridization to determine the percentage of cells actively transcribing the transgene (Wijgerde *et al.*, 1995). Probes detecting transcription of the mouse α-globin genes were used as a control to establish the number of erythroid cells in each preparation. Figure 1D shows that the β-globin transgene is actively transcribed in >95% of the cells in EKLF$^{+/+}$ E10.5 blood (Table I). In the absence of EKLF, active transcription of the human β-globin gene is observed in ~55% of the cells, in good agreement with the S1 nuclease analysis. Interestingly, human β-globin gene transcription is detected in <1% of the cells in EKLF$^{-/-}$ fetal liver, while 95% of E13.5 EKLF$^{+/+}$ fetal liver cells actively transcribe the transgene (Figure 1D and Table I). The extremely low number of cells expressing the transgene in EKLF$^{-/-}$ mice closely resembles previous observations on adult β-globin gene transcription (Wijgerde *et al.*, 1996). Thus, the relatively high level of human β-globin mRNA in E13.5 fetal liver of EKLF$^{-/-}$ mice observed by S1 nuclease analysis is due to earlier expression of the transgene in primitive erythroid cells still present in the circulation at E13.5.

We conclude that EKLF has only a modest effect on the expression of the human β-globin transgene in primitive cells, implying that expression of the β-globin gene is not intrinsically dependent on EKLF in this lineage. However, EKLF is essential for β-globin expression in definitive cells, regardless of the presence of embryonic and fetal globin genes.

### *EKLF is a transcriptional activator in primitive erythroid cells required for the function of 5'HS3*

The above results gave the first indication that EKLF could function as a transcriptional activator in the primitive lineage, since we observed a reduced expression of the β-globin transgene in EKLF$^{-/-}$ primitive cells. To further study the potential role of EKLF in LCR-mediated gene activation in embryonic erythroid cells, we used *lacZ* reporter mice in which μLCR constructs drive expression of the *lacZ* gene from a basic, non-erythroid promoter (Tewari *et al.*, 1996). Three different constructs were used (Figure 2A). The first contained the μLCR, as used in the μLCR–β-globin construct described above. The second construct contained the 1.9 kb 5'HS3 fragment of the μLCR, and the third construct contained only the 225 bp core fragment of 5'HS3 with six potential EKLF binding sites (Philipsen *et al.*, 1993; Philipsen *et al.*, 1990). *LacZ* expression of the μLCR and 5'HS3 transgenes is confined to primitive cells, while the core 5'HS3 transgene is also expressed in definitive erythroid cells (not shown; Tewari *et al.*, 1996).

The three reporter lines were bred into the EKLF$^{-/-}$ background, and expression of the *lacZ* reporter gene in primitive cells was assayed by X-gal staining of E10.5

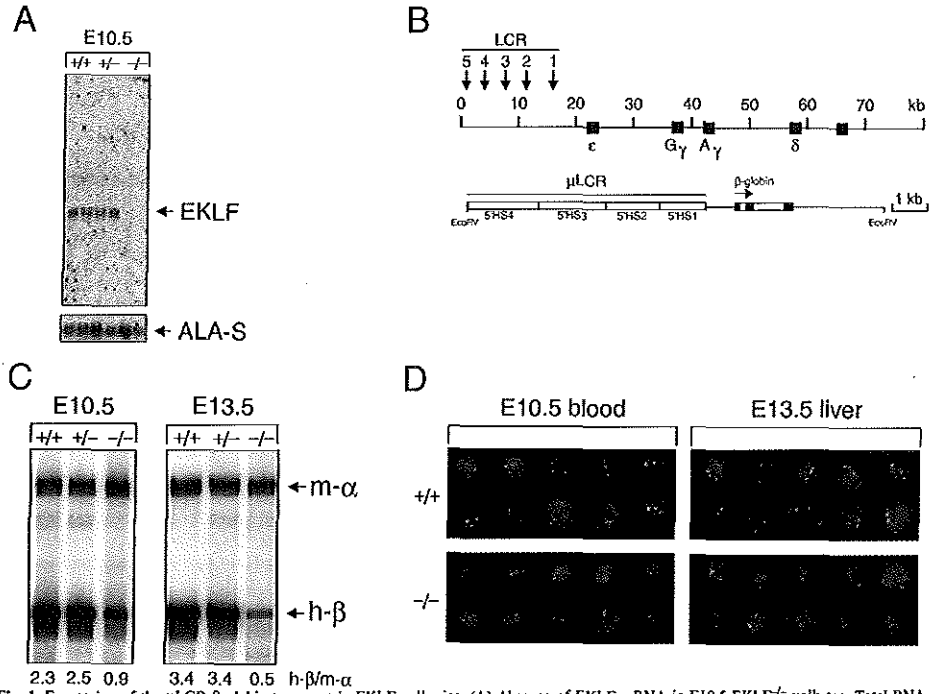



**Fig. 1.** Expression of the μLCR-β-globin transgene in EKLF null mice. (A) Absence of EKLF mRNA in E10.5 EKLF$^{-/-}$ yolk sac. Total RNA was isolated from E10.5 day yolk sacs, subjected to Northern blotting and hybridized with EKLF cDNA to demonstrate the absence of EKLF encoding mRNA in the knock-out samples (top). To control for loading of the gel, the same blot was rehybridized with a cDNA encoding the haem synthesis enzyme ALA-S (bottom). (B) The human β-globin locus and the μLCR-β-globin construct. The μLCR was directly linked to the human β-globin gene, and this construct was used to generate a single copy transgenic mouse line (μD14 in Ellis *et al.*, 1996). (C) The expression of the μLCR-β-globin transgene in primitive and definitive cells of +/+, +/- and -/- EKLF fetuses. The fetuses were dissected at the stages indicated, genotyped, and RNA was isolated from the whole embryo (E10.5) or the fetal liver (E13.5). Expression of the transgene was determined with the S1-nuclease protection assay; the endogenous mouse α-globin mRNA was used as an internal control. The results were quantitated on a PhosphorImager. The specific activity of the human β-globin probe was approximately six times higher than that of the mouse α-globin probe. (D) Primary transcript *in situ* hybridization of transgene expression in EKLF knock-out mice. Cells were isolated from E10.5 blood or E13.5 fetal liver and subjected to *in situ* hybridization to detect active transcription of the β-globin transgene (Wijgerde *et al.*, 1995). Genotypes of the fetuses are as indicated. Representative examples of cells with signals corresponding to α-globin (green) and human β-globin (red) are shown.

**Table I.** *In situ* hybridization analysis of human β-globin transcription in E10.5 blood and E13.5 liver cells of EKLF knock-out mice

| Genotype | Time point | Percentage of cells with human β-globin nuclear signal | Percentage of cells with α-globin nuclear signals |
|---|---|---|---|
| μLCR/β-globin, EKLF$^{+/+}$ | E10.5 | 98 | 98 |
| μLCR/β-globin, EKLF$^{-/-}$ | E10.5 | 56 | 94 |
| μLCR/β-globin, EKLF$^{+/+}$ | E13.5 | 95 | 98 |
| μLCR/β-globin, EKLF$^{-/-}$ | E13.5 | <1 | 96 |

embryos (Bonnerot and Nicolas, 1993). At E10.5, erythroid expression of the μLCR-*lacZ* transgene is easily detectable in EKLF$^{-/-}$ mice (Figure 2B). X-gal staining of erythroid cells is greatly reduced in the (multi-copy) 5'HS3-*lacZ* line (not shown) and undetectable in the (single copy) core 5'HS3-*lacZ* line (Figure 2C); this is particularly evident when the yolk sacs are stained (compare with Figure 2D and E). Note that the position effect, i.e. *lacZ* expression outside the erythroid system (Tewari *et al.*, 1996), is unaffected in the core 5'HS3-*lacZ*/EKLF$^{-/-}$ embryos (Figure 2C). To demonstrate that the observed effects on *lacZ* activity are due to reduced transcription of the transgenes, we performed *in situ* hybridizations to determine the percentage of cells actively transcribing the *lacZ* reporter gene (Wijgerde *et al.*, 1995; Tewari *et al.*, 1996). These experiments are summarized in Table II and show that there is a small reduction (from 14 to 10%) in the number of cells with *lacZ* transcription spots in μLCR-

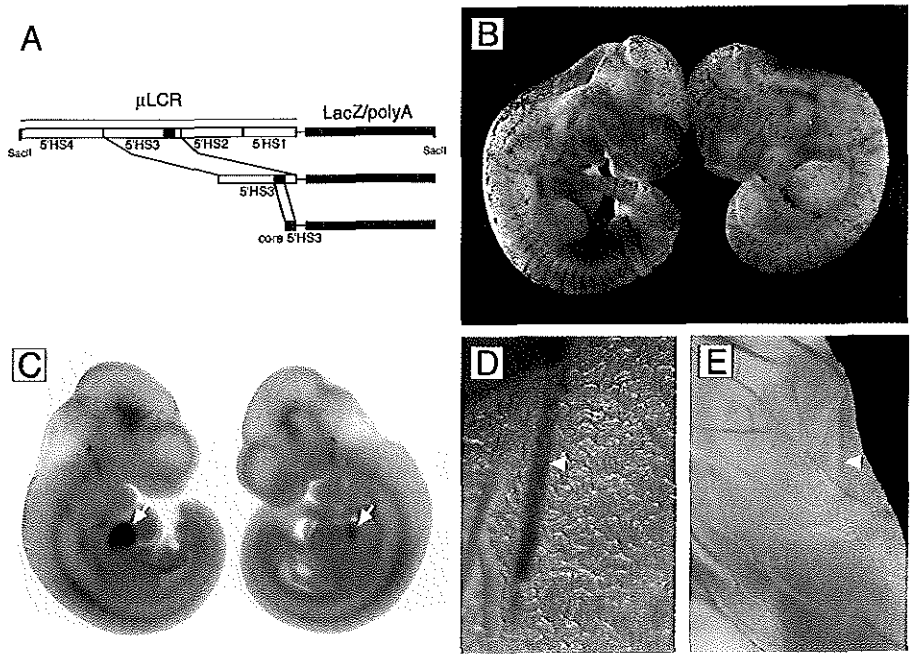



**Fig. 2.** Activation of μLCR- and 5'HS3-*lacZ* reporter genes in EKLF knock-out mice. (A) Schematic drawing of the transgene constructs. The multi-copy μLCR-*lacZ* and 5'HS3-*lacZ* transgenic lines have been described (lines 'b' in Tewari *et al.*, 1996), the core 5'HS3-*lacZ* line carries a single copy of the transgene. (B) Embryos were dissected at E10.5, fixed and stained for β-galactosidase activity as described previously (Bonnerot and Nicolas, 1993; Tewari *et al.*, 1996). Shown are a μLCR-*lacZ*/EKLF$^{+/+}$ (left) and a μLCR-*lacZ*/EKLF$^{-/-}$ (right) embryo. (C) As (B), except that core 5'HS3-*lacZ*/EKLF$^{+/+}$ (left) and core 5'HS3-*lacZ*/EKLF$^{-/-}$ (right) embryos are shown. Arrows indicate the presence of embryonic blood in the heart which does not stain blue in the core 5'HS3-*lacZ*/EKLF$^{-/-}$ embryo. Note that the position effect of the core 5'HS3-*lacZ* line is unaffected by the EKLF knock-out (see text). (D) X-gal stained yolk sacs of E10.5 core 5'HS3-*lacZ*/EKLF$^{+/+}$ and (E) core 5'HS3-*lacZ*/EKLF$^{-/-}$ embryos, showing the absence of β-galactosidase activity in circulating EKLF$^{-/-}$ primitive cells; arrowheads indicate blood-filled arteries.

*lacZ*/EKLF$^{-/-}$ transgenics. In 5'HS3-*lacZ*/EKLF$^{-/-}$ and core 5'HS3/EKLF$^{-/-}$ mice, there is a >10-fold reduction in the number of cells with *lacZ* transcription spots; a >10-fold reduction was also found in E13.5 fetal liver cells of core 5'HS3/EKLF$^{-/-}$ mice, as expected. Together, these data imply that EKLF is not essential for transcriptional activation by the μLCR, in agreement with observations on the μLCR-β-globin transgene (this paper) and the complete human β-globin locus (Perkins *et al.*, 1996; Wijgerde *et al.*, 1996). However, the activity of 5'HS3 and the core of 5'HS3 is dependent on the presence of EKLF. Furthermore, our data show unambiguously that EKLF is a functional transcriptional activator in primitive cells. This activity of EKLF has previously gone unnoticed (Nuez *et al.*, 1995; Perkins *et al.*, 1995; Wijgerde *et al.*, 1996).

### *Increased EKLF levels expedite the switch from γ- to β-globin expression*

It has been proposed that globin switching is mediated through a gradual change in the transcription factor environment (Grosveld *et al.*, 1993), and previous experiments carried out with transgenic mice harbouring the complete human β-globin locus showed that the switch from γ- to β-globin expression, which occurs in the fetal liver between E12 and E14, is delayed in EKLF$^{+/-}$ mice (Wijgerde *et al.*, 1996). This suggests that EKLF is present at critical levels in switching cells, thus we asked if increased EKLF levels would also influence switching. To investigate this, we cloned a cDNA encoding the mouse EKLF protein in the expression vector pEV3 (Needham *et al.*, 1992). This vector contains the μLCR, the first 400 bp of the β-globin promoter, a modified β-globin gene and the 3' flanking region of the β-globin gene (Figure 3A). Two transgenic mouse lines (EKLF-

**Table II.** *In situ* hybridization analysis of *lacZ* transcription in E10.5 blood of EKLF knock-out mice

| Genotype | Percentage of cells with lacZ nuclear signal | Percentage of cells with α-globin nuclear signals |
|---|---|---|
| μLCR-*lacZ*, EKLF$^{+/+}$ | 14 | 95 |
| μLCR-*lacZ*, EKLF$^{-/-}$ | 10 | 92 |
| 5'HS3-*lacZ*, EKLF$^{+/+}$ | 15 | 97 |
| 5'HS3-*lacZ*, EKLF$^{-/-}$ | 1 | 97 |
| core5'HS3-*lacZ*, EKLF$^{+/+}$ | 19 | 98 |
| core5'HS3-*lacZ*, EKLF$^{-/-}$ | <1 | 92 |

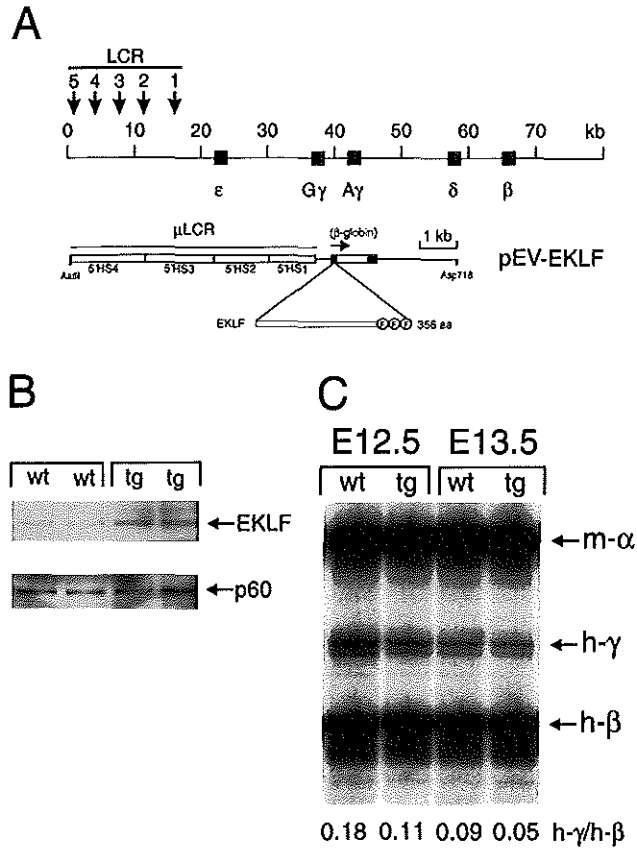



**Fig. 3.** Expression of EKLF under the control of the β-globin promoter and LCR. **(A)** The cDNA encoding EKLF was cloned in the expression vector pEV3 (Needham *et al.*, 1992). The LCR and a modified version of the β-globin gene, used to drive expression of EKLF cDNA in erythroid cells, are shown. Restriction enzyme sites used to isolate the microinjection fragment for transgenesis are indicated. **(B)** Western blot analysis of EKLF expression. Ten μg of protein, isolated from E13.5 fetal livers with the indicated genotype (wt, wild-type; tg, pEV-EKLF transgenic), were fractionated on a 12.5% PAA–SDS gel and blotted onto a PVDF membrane. The blot was probed with a rabbit anti-EKLF polyclonal antibody (top). As a loading control, the blot was reprobed with a mouse monoclonal antibody recognizing the 62 kDa subunit of TF-IIH. **(C)** The effect of increased EKLF levels on γ- and β-globin gene expression. Mice carrying the complete 70 kb human β-globin locus were crossed with the EKLF-pEV transgenics. Fetuses were dissected at the stages indicated, genotyped, and the expression of human γ- and β-globin mRNA was determined with quantitative S1-nuclease protection analysis (Wijgerde *et al.*, 1996). The reduction in the γ:β ratio in the presence of the pEV-EKLF transgene was consistently found in three independent litters for each time point (data not shown). The probes were labelled to approximately equal specific activities.

pEV) were generated with this construct. Western blot analysis of E13.5 fetal liver protein shows modest over-expression (2- to 3-fold) of EKLF in one of these lines (Figure 3B). This line was crossed with mice containing the human β-globin locus (Strouboulis *et al.*, 1992) to analyse the expression of γ- and β-globin mRNA in E12.5 and E13.5 fetal liver. The results of the S1 nuclease analysis are shown in Figure 3C. The level of endogenous mouse α-globin mRNA was used to control for the amount of RNA used in each sample. Quantitation of these data reveals that the ratio of γ/β-globin expression decreases earlier in mice carrying the EKLF-pEV transgene, i.e. with a higher EKLF level. Thus, we conclude that the rate of γ- to β-globin gene switching correlates with the level of EKLF (Wijgerde *et al.*, 1996).

### *The EKLF transgene rescues the lethal phenotype of the EKLF null mutation*

The EKLF-pEV mice should express transgene-derived EKLF in primitive and definitive cells (Hanscombe *et al.*, 1991). Thus, this transgene might rescue the lethality of the EKLF null mutation. To test this, we bred the EKLF-pEV transgene into the EKLF$^{-/-}$ background. In the F2 generation, we obtained mice with an EKLF-pEV/EKLF$^{-/-}$ genotype with both transgenic lines at the expected Mendelian ratio (data not shown). As could be expected,

Table III. Haematological parameters of adult pEV-EKLF rescue mice

| | EKLF +/+ ($n$ = 7) | EKLF +/– ($n$ = 7) | EKLF–/– pEV-EKLF ($n$ = 7) | EKLF+/+ pEV-EKLF ($n$ = 6) |
|---|---|---|---|---|
| RBC ($\times 10^{12}$) | 8.8 ± 0.5 | 8.3 ± 0.5 | 6.5[a] ± 0.8 | 8.9 ± 0.6 |
| HGB (mmol/l) | 8.2 ± 0.3 | 7.6[b] ± 0.5 | 6.4[a] ± 0.8 | 8.0 ± 0.5 |
| HCT (l/l) | 0.51 ± 0.02 | 0.49 ± 0.03 | 0.41[a] ± 0.05 | 0.51 ± 0.03 |
| MCV (fl) | 58.4 ± 2.4 | 59.6 ± 1.5 | 63.7[a] ± 3.4 | 57.4 ± 1.5 |
| PLT ($\times 10^9$) | 1360 ± 63 | 1391 ± 138 | 1213[b] ± 139 | 963[a] ± 50 |

RBC, red blood cells; HGB, haemoglobin; HCT, haematocrit; MCV, mean cell volume; PLT, platelets.
[a]indicates values that are very significantly different from the wild-type ($p$<0.004) using a Student's or Welch unpaired $t$-test. Data are given ± SD.
[b]indicates values that are significantly different from the wild-type (0.04>$p$>0.004).

the EKLF-pEV transgene also rescues expression in the primitive cells of the 5'HS3-*lacZ* mice discussed above (data not shown).

Haematological analysis of adult mice shows that EKLF-pEV/EKLF$^{-/-}$ mice have lower red blood cell counts, haemoglobin content and haematocrit in comparison with wild-type mice, while the mean cell volume of the erythrocytes is increased (Table III). Thus, erythropoiesis is still slightly impeded in the rescue mice, which might be caused by inadequate expression of transgene-derived EKLF. In addition, we observe that platelet counts are lower in EKLF-pEV/EKLF$^{-/-}$, and even further reduced in EKLF-pEV/EKLF$^{+/+}$ mice. Finally, in agreement with the analysis of fetal erythropoiesis (Nuez *et al.*, 1995; Perkins *et al.*, 1995), the EKLF$^{+/-}$ mice have haematological values that are close to those found for EKLF$^{+/+}$ mice, although the slightly reduced haemoglobin content might indicate a very mild phenotype.

In conclusion, we show that the expression of transgene-derived EKLF rescues the lethal phenotype of the EKLF null mutation. This formally proves that the phenotype of the EKLF$^{-/-}$ fetuses was indeed the result of the mutation in the EKLF gene, and not some other defect in the ES cells or an unknown other gene that might be present at the EKLF locus.

## Discussion

### *Expression of the β-globin gene is not completely dependent on EKLF in primitive cells*

EKLF is essential for high level expression of the adult β-globin genes, and EKLF$^{-/-}$ mice die of anaemia at E14–15 due to defective erythropoiesis caused by severely reduced β-globin levels (Nuez *et al.*, 1995; Perkins *et al.*, 1995). Transgenic mice with just a β-globin gene driven by the μLCR express the transgene prematurely in primitive cells (Hanscombe *et al.*, 1991); we have used a single copy transgenic line with such a construct (Ellis *et al.*, 1996) to assess the potential role of EKLF in embryonic β-globin expression. Surprisingly, there is only a modest reduction in human β-globin mRNA levels in E10.5 EKLF$^{-/-}$ embryos. This reduction is reflected in the number of cells actively transcribing the transgene, as analysed by primary transcript *in situ* hybridization (Wijgerde *et al.*, 1995). In EKLF$^{+/+}$ cells, transcription foci of the transgene are detected in the nuclei of >95% of the erythroid cells, and in EKLF$^{-/-}$ cells in ~55% of the nuclei. This indicates that the β-globin gene is transcribed only part of the time in –/– primitive cells, which could be attributed to a less stable interaction between the LCR and the β-globin promoter (Milot *et al.*, 1996). However, the gene is transcriptionally competent and still transcribed at high levels. This is in sharp contrast with the observations in definitive cells of E13.5 EKLF$^{-/-}$ fetal liver. Primary transcript *in situ* hybridization shows a dramatic reduction in the number of cells actively transcribing the β-globin gene. In agreement with data on the expression of the β-globin gene in the intact globin locus (Wijgerde *et al.*, 1996), fewer than 1% of the cells contain a nuclear transcription focus of the transgene. In control littermates, a transcription signal is detectable in >95% of the cells. Thus, expression of the β-globin gene is strictly dependent on the presence of EKLF in definitive cells. However, the expression in primitive cells shows that the β-globin gene is not intrinsically dependent on EKLF. Rather, this dependency is acquired in definitive cells through an as yet unidentified mechanism. A more restrictive chromatin environment in definitive cells, as has been proposed to explain the embryonic expression pattern of many μLCR-*lacZ* transgenes, could be such a mechanism (Tewari *et al.*, 1996). Consequently, the autonomous silencing of the ε- and γ-globin genes (Grosveld *et al.*, 1993) might at least in part be an intrinsic property of these genes mediated by repressive chromatin interactions.

### *EKLF is required for the function of 5'HS3 and is a transcriptional activator in primitive erythropoiesis*

The promoter of the human β-globin gene contains a canonical binding site for EKLF (Feng *et al.*, 1994). Thus, it is likely that direct binding of EKLF to this element is mandatory for high level expression of the gene in definitive cells (Perkins *et al.*, 1996; Wijgerde *et al.*, 1996). In addition, EKLF binding sites are found in the hypersensitive sites of the LCR (Grosveld *et al.*, 1993). The best example is provided by 5'HS3, the only element of the LCR that can reproducibly direct expression of single copy globin transgenes in mice and, therefore, has been postulated to provide a dominant chromatin opening activity (Ellis *et al.*, 1996). The 225 bp core fragment of 5'HS3 contains six potential EKLF binding sites (Philipsen *et al.*, 1990). Intactness of these binding sites is essential for the activity of 5'HS3 (Philipsen *et al.*, 1993). Due to the overlapping binding specificity of proteins present in erythroid nuclei, most notably those belonging to the Sp1 family of transcription factors, it is not known which proteins bind to these sequences *in vivo* (Philipsen *et al.*, 1993). To separate the LCR binding sites from those of the promoter, we used mice with a *lacZ* reporter gene

and a non-erythroid minimal promoter linked to LCR derivatives (Tewari *et al.*, 1996) and analysed reporter gene expression in EKLF null mutants. Consistent with the analyses of the β-globin loci (Perkins *et al.*, 1996; Wijgerde *et al.*, 1996), the μLCR-*lacZ* construct is still active in EKLF$^{-/-}$ embryos. In contrast, expression of the 5'HS3-*lacZ* and core 5'HS3-*lacZ* transgenes is greatly reduced in E10.5 blood, showing that EKLF is required for transcriptional activation by 5'HS3. These results suggest that EKLF binds directly to the core fragment of 5'HS3. Our observations are consistent with the reduced hypersensitivity of 5'HS3 in EKLF$^{-/-}$ fetal livers (Wijgerde *et al.*, 1996) and the phenotype of the 5'HS3 deletion in mice (Hug *et al.*, 1996). This deletion has a very small effect on the expression of the linked embryonic εy and βh1 genes, while adult β-globin expression is reduced by 30%.

The results with the *lacZ* transgenics show that EKLF is a functional transcriptional activator in primitive cells, in addition to its previously described activity in definitive cells (Nuez *et al.*, 1995; Perkins *et al.*, 1995; Wijgerde *et al.*, 1996). Thus, EKLF may activate specific genes in the primitive lineage, but reduced expression of these genes in EKLF$^{-/-}$ embryos would not result in an obvious phenotype. Experiments aimed at the identification of such EKLF target genes in the primitive and definitive lineage are in progress.

### *The level of EKLF influences globin gene switching*

We have assessed the effect of increased EKLF levels on γ- to β-globin switching in transgenic mice with the complete human β-globin locus (Strouboulis *et al.*, 1992). Globin switching is a gradual process in which the cells change from γ- to β expression over a period of ~2 days (Wijgerde *et al.*, 1995). It has been proposed that switching is effectuated by subtle changes in transcription factor environment (Grosveld *et al.*, 1993). In accord with this notion, our data show that higher EKLF levels result in an earlier switch from γ- to β-globin expression in the fetal liver. Thus, the timing of switching is influenced by EKLF levels, and higher than normal EKLF levels expedite the switch. These data are compatible with the prevailing model that the LCR acts as a holocomplex activating only one gene at a time (Wijgerde *et al.*, 1995). EKLF might increase the stability of interactions between the LCR and the β-globin promoter. Hence, we propose that higher EKLF levels will reduce the time the LCR spends activating the γ-globin genes, thus resulting in an earlier switch (Wijgerde *et al.*, 1996). After switching has completed, increased EKLF levels no longer affect β-globin gene expression, presumably because EKLF is present at saturating levels in adult erythroid cells. This is consistent with the observation that the mouse and human β-globin genes are expressed at wild-type levels in EKLF$^{+/-}$ mice (Wijgerde *et al.*, 1996).

### *The EKLF-pEV transgene rescues the lethal phenotype of EKLF$^{-/-}$ mice*

We show that the lethal phenotype of EKLF$^{-/-}$ mice can be rescued by expressing EKLF under the control of the β-globin LCR. The haematological analysis shows that, compared with wild-type mice, adult EKLF rescue mice have lower red blood cell numbers, haemoglobin concentration and haematocrit, and a slightly increased mean cell volume. Thus, a subtle phenotype remains in these rescue mice which indicates that appropriate developmental expression of EKLF is important for normal erythropoiesis. This notion is emphasized by the observation that pEV-EKLF/EKLF$^{+/+}$ mice have significantly reduced numbers of platelets, suggesting that EKLF may affect the balance between the megakaryocytic and erythroid lineages. Hence, there might be an unexpected link between EKLF and megakaryopoiesis, as has recently been described for another erythroid-specific transcription factor, GATA-1 (Shivdasani *et al.*, 1997).

The rescue experiment sets a platform for the analysis of EKLF protein domains required for high level β-globin expression. EKLF transactivation domains have been characterized in cell culture assays, and a cellular factor interacting with EKLF has been implicated in EKLF-mediated transcativation (Chen and Bieker, 1996). These observations can now be stringently tested through the rescue of EKLF$^{-/-}$ mice. Furthermore, co-immunoprecipitation of EKLF and its interaction partners from primary erythroid cells could be attempted after the introduction of an epitope tag into the EKLF transgene.

## Materials and methods

#### *Transgenic mice*

The EKLF-pEV construct was generated by inserting the EKLF cDNA into the *Bgl*II site of the pEV3 expression vector (Needham *et al.*, 1992). The microinjection fragment was released by digestion with *Aat*II and *Asp*718, gel purified and used for transgenesis as described (Kollias *et al.*, 1986). The 225 bp core fragment of 5'HS3 (Philipsen *et al.*, 1990) was cloned in a *lacZ* reporter vector (Kothary *et al.*, 1988) and digested with *Hin*dIII and *Asp*718 to obtain the microinjection fragment. The other transgenic lines used in this study have been described before: μLCR-β-globin line μD 14 (Ellis *et al.*, 1996); μLCR-*lacZ* line b and 5'HS3-*lacZ* line b (Tewari *et al.*, 1996); 70 kb human β-globin locus line 72 (Strouboulis *et al.*, 1992) and EKLF knock-out mice (Nuez *et al.*, 1995).

#### *DNA and RNA analysis*

Genomic mouse DNA was prepared from part of the body of the fetuses or tail clips from adult mice. The mice were genotyped by Southern blot analysis (Strouboulis *et al.*, 1992; Nuez *et al.*, 1995; Ellis *et al.*, 1996; Tewari *et al.*, 1996). RNA was isolated from erythroid tissues and globin gene expression was quantitated by S1 nuclease analysis as described (Strouboulis *et al.*, 1992; Wijgerde *et al.*, 1996). Northern blotting was done as described (Talbot *et al.*, 1989). Probes used were the complete EKLF cDNA and a cDNA detecting the erythroid-specific form of aminolevulinate synthase (ALA-S), a kind gift of Dr Peter Curtis (Boston).

#### *Protein analysis*

Protein was isolated from E13.5 liver, and 10 μg was used for Western blot analysis with a rabbit anti-EKLF polyclonal antibody (Southwood *et al.*, 1996), a kind gift of Dr J.J.Bieker (New York). A mouse monoclonal antibody recognizing the 62 kDa subunit of THII-H was used as a loading control; this antibody was generously provided by Drs B.Winkler and G.Weeda (Rotterdam).

#### *LacZ staining*

The day of the vaginal plug was taken as E0.5. The embryos were dissected out of the decidua and fixed on ice in a solution containing 1% formaldehyde and 0.5% glutaraldehyde in PBS for 30 min. They were then rinsed in embryo buffer (PBS, 0.02% NP40, 0.01% deoxycholate, 2 mM $MgCl_2$) and incubated in staining solution [1 mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$ and 1 mM EGTA in embryo buffer]. After 3–4 h at 37°C, or overnight at room temperature in the dark, the staining reaction was stopped by removing this solution and rinsing the embryos in PBS.

### *In situ hybridization on erythroid cells*

E10.5 embryonic blood from the heart or E13.5 liver cells from individual fetuses were disrupted in 50 and 150 µl of PBS respectively. Twenty µl of this cell suspension was fixed onto a poly-lysine coated slide in 4% formaldehyde, 5% acetic acid for 20 min at room temperature. The cells were then washed three times for 10 min in PBS and stored in 70% ethanol at –20°C.

Three probes were utilized for *in situ* hybridization. A nick-translated biotinylated *lacZ* probe was used for the detection of *lacZ* transcription (Tewari *et al.*, 1996). A set of four biotinylated oligonucleotides was used to detect human β-globin transcription (Dillon *et al.*, 1997). DIG-labelled mouse α-globin intron-specific oligonucleotides were used to reveal the erythroid cells (Tewari *et al.*, 1996). Hybridizations were done as described (Wijgerde *et al.*, 1995; Tewari *et al.*, 1996). The *lacZ* hybridization mix contained 50% formamide and 10% dextran sulfate. For antibody detection, the Tyramid amplification system (TSA) of Dupont was used (Raap *et al.*, 1995). For quantitation, ~300 cells were scored for each slide. The reproducibility of these data fell within a 10% error margin.

### *Haematological analysis*

Adult mice (3–5 months of age) were bled by orbital bleeding, and blood samples were analysed on an F800 microcell counter. Statistical analysis was done by unpaired two-sided *t*-testing.

## Acknowledgements

We thank Lien Braam for animal husbandry. R.T. and M.W. were supported by an N.W.O. (The Netherlands) grant to F.G.

## References

Bonnerot,C. and Nicolas,J.-F. (1993) Application of *lacZ* gene fusions to postimplantation development. *Methods Enzymol.*, **225**, 451–469.

Chen,X. and Bieker,J.J. (1996) Erythroid Krüppel-like factor (EKLF) contains a multifunctional transcriptional activation domain important for inter- and intramolecular interactions. *EMBO J.*, **15**, 5888–5896.

Dillon,N., Trimborn,T., Strouboulis,J., Fraser,P. and Grosveld,F. (1997) The effect of distance on long-range chromatin interactions. *Molecular Cell*, **1**, 131–139.

Dzierzak,E. and Medvinsky,A. (1995) Mouse embryonic hematopoiesis. *Trends Genet*, **11**, 359–366.

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

Feng,W.C., Southwood,C.M. and Bieker,J.J. (1994) Analyses of β-thalassemia mutant DNA interactions with erythroid Krüppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J. Biol. Chem.*, **269**, 1493–1500.

Grosveld,F., Dillon,N. and Higgs,D. (1993) The regulation of human globin gene expression. *Baillieres Clin. Haematol.*, **6**, 31–55.

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

Hug,B.A., Wesselschmidt,R.L., Fiering,S., Bender,M.A., Epner,E., Groudine,M. and Ley,T.J. (1996) Analysis of mice containing a targeted deletion of β-globin locus control region 5′ hypersensitive site 3. *Mol. Cell. Biol.*, **16**, 2906–2912.

Kollias,G., Wrighton,N., Hurst,J. and Grosveld,F. (1986) Regulated expression of human A γ-, β-, and hybrid γβ-globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell*, **46**, 89–94.

Kothary,R., Clapoff,S., Brown,A., Campbell,R., Peterson,A. and Rossant,J. (1988) A transgene containing *lacZ* inserted into the dystonia locus is expressed in neural tube. *Nature*, **335**, 435–437.

Miller,I.J. and Bieker,J.J. (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol. Cell. Biol.*, **13**, 2776–2786.

Milot,E. *et al.* (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell*, **87**, 105–114.

Needham,M., Gooding,C., Hudson,K., Antoniou,M., Grosveld,F. and Hollis,M. (1992) LCR/MEL: a versatile system for high-level expression of heterologous proteins in erythroid cells. *Nucleic Acids Res.*, **20**, 997–1003.

Nuez,B., Michalovich,D., Bygrave,A., Ploemacher,R. and Grosveld,F. (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, **375**, 316–318.

Perkins,A.C., Sharpe,A.H. and Orkin,S.H. (1995) Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, **375**, 318–322.

Perkins,A.C., Gaensler,K.M. and 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.

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

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

Raap,A.K., van de Corput,M.P., Vervenne,R.A., van Gijlswijk,R.P., Tanke,H.J. and Wiegant,J. (1995) Ultra-sensitive FISH using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. *Hum. Mol. Genet.*, **4**, 529–534.

Shivdasani,R.A., Fujiwara,Y., McDevitt,M.A. and Orkin,S.H. (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.

Southwood,C.M., Downs,K.M. and Bieker,J.J. (1996) Erythroid Krüppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev. Dyn.*, **206**, 248–259.

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

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

Tewari,R., Gillemans,N., Harper,A., Wijgerde,M., Zafarana,G., Drabek,D., Grosveld,F. and Philipsen,S. (1996) The human β-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial *lacZ* gene. *Development*, **122**, 3991–3999.

Thein,S.L. (1993) β-thalassaemia. *Baillieres Clin. Haematol.*, **6**, 151–175.

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

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

*Received November 17, 1997; revised January 5, 1998; accepted February 2, 1998*

# Chapter 5

## Chromatin interaction mechanism of transcriptional control

*Gribnau, J., de Boer, E., Trimborn, T., Wijgerde, M., Grosveld, F. and Fraser, P.*

# Chromatin interaction mechanism of transcriptional control

Joost Gribnau, Ernie de Boer, Tolleiv Trimborn, Mark Wijgerde, Eric Milot, Frank Grosveld and Peter Fraser

*MGC Department of Cell Biology and Genetics, Erasmus University, PO Box 1738, 3000 DR Rotterdam, Netherlands.*

## Abstract

**We have used a kinetic analysis to distinguish possible mechanisms of activation of transcription of the different genes in the β globin locus. Based on *in situ* studies at single cell level we have previously suggested a dynamic mechanism of single genes alternately interacting with the Locus Control Region (LCR). However those steady state experiments did not allow a direct measurement of the dynamics of the mechanism and the presence of loci with two gene signals *in cis* has left open the possibility that multiple genes in the locus could initiate transcription simultaneously. By utilizing transcription inhibition and release in conjunction with RNA FISH we show that multiple β-gene primary transcript signals *in cis* represent a transition between alternating transcriptional periods of single genes, supporting a dynamic interaction mechanism.**

## Introduction

Several models have been proposed for the activation of gene transcription after chromatin activation of a locus. In one model, the genes would simply be accessible to binding of transcription factors and be transcribed in a stochastic fashion (Groudine and Weintraub, 1982; Martin et al., 1996). In such a model the dimensional aspects of the locus would not play an important role. In a second model regulatory sequences could be the entry site of (part of) the transcriptional machinery which would scan the DNA for genes to be transcribed (Herendeen et al., 1992; Tuan et al., 1992). Such a model is basically linear and predicts that the order of the genes relative to the regulator is an important parameter. In a third model it has been proposed that gene regulatory elements participate in direct chromatin interactions with regulatory elements at a large distance as a prerequisite to transcriptional activation (Ptashne, 1988; Bickel and Pirotta, 1990; Mueller-Storm et al., 1989; Foley and Engel, 1992; Wijgerde et al., 1995; Dillon et al., 1997). Such a looping model is three dimensional and predicts that the relative distance of the genes to the regulator is important. All these models have been put forward to explain the role of the human β-globin locus control region (LCR) in the developmental regulation of transcription of the β-gene cluster (Martin et al., 1996; Tuan et al., 1992; Wijgerde et al., 1995; Dillon et al., 1997).

The β-globin system has long been a prototypic system for the study of transcription in vertebrates (reviewed in Grosveld et al., 1993). The locus consists of five active genes that are activated and silenced at different stages of erythroid development (Fig 1). The expression of all of these genes is dependent on the presence of the LCR which is located 15kb upstream from the ε gene (Grosveld et al., 1987). The ε gene is expressed first in the embryonic yolk sac followed by a gradual switch to expression of the γ genes between six and ten weeks of gestation. They predominate during the fetal liver stage. In the late fetal liver and neonatal stages, there is a second transition to expression of the β gene and the γ genes are almost completely silenced during adult life. When the locus is introduced in transgenic mice a similar expression pattern is observed, although the γ genes are expressed early in the embryo and are switched off at day 16 of development in the fetal liver (Strouboulis et al., 1992; Peterson et al., 1993). The analysis of mutated loci found in patients and the use of single ε, γ and β genes in transgenic mice have shown that the ε and γ genes are supressed autonomously through sequences directly flanking the genes (Dillon et al., 1991; Raich et al., 1990). However the β globin gene when present in the whole locus is (at least in large part) silenced during early development in a non autonomous manner. This β-globin gene suppression can be explained by a scanning mechanism because genes closer to the regulatory sequences would have a natural advantage over distal genes due to proximity. This would also be the case in a looping mechanism, proximal genes would have a higher frequency of interaction with the regulatory sequences and thus have a competitive advantage over distal genes (Giglioni et al., 1984; Enver et al., 1990; Hanscombe et al., 1991; Peterson and Stamatoyannopoulos, 1993; Dillon et al., 1997). However this would not be the case in an accessability model and hence an extra parameter was postulated to explain the silencing of the distal β-gene in early development, namely a process of interference of the proximal genes with the distal genes via some topological constraint (Martin et al., 1996).

Recent analysis of primary transcription in single cells (Wijgerde et al., 1995, 1996) and the results obtained by placing a second β globin gene at different positions in the locus (Dillon et al., 1997) support a dynamic looping mechanism with single genes alternately interacting with the LCR. However, the presence of a minority of loci which display two gene signals *in cis* (Wijgerde et al., 1995) could be interpreted as evidence in support of the scanning or the accessability model of initiation (Martin et al 1996). Thus a crucial difference between the looping model and the others is single vs multiple gene activation at any moment in any locus. We have therefore used a novel kinetic analysis utilizing transcription inhibition and release in conjunction with RNA FISH to show that multiple β-gene primary transcript signals *in cis* represent a transition between alternating transcriptional periods of single genes,

rather than the co-initiation of transcription of multiple genes in the locus.

## Results

### The lifetime of the *in situ* hybridization signals

A key parameter in a kinetic analysis of the transcription process *in vivo* using *in situ* hybridization, is the time required for a signal to decay below the level of detection. To enable the detection of short lived events we probed for the presence of intronic RNA as these sequences are cleaved rapidly from the primary transcript and degraded. We used actinomycin-D to measure the detection lifetime of the primary transcript signals at days 11.5 (Fig. 1) and 12.5 (not shown) of development in the foetal liver. On both day 11.5 and 12.5, the intron signals of the ζ, α, γ and β-genes all disappear below detection level 7.5 minutes after the addition of actinomycin-D (Fig. 1c and d). Interestingly the intron signals of the ζ and the α genes do not decrease immediately when compared to the β-like genes. This could indicate a possible difference between the two loci (see below), but could also be due to experimental parameters such as a higher sensitivity of the α-like probes. Unfortunately actinomycin-D inhibition is irreversible and hence cannot be used in reactivation experiments.

We therefore used 5,6-dichloro-1-b-D-ribofuranosylbenzimidazole (DRB) to reversibly inhibit pol-II transcriptional elongation in single copy human β-globin transgenic mouse (Strouboulis et al., 1992) fetal liver cells. Previous studies have shown that DRB does not affect initiation of transcription (Fraser et al., 1978; Marshall et al., 1992) but prematurely aborts elongating transcripts up to 400-600 bp from the initiation site (Chodosh et al., 1989; Marshall et al., 1996) by inhibiting the activity of the P-TEFb kinase which phosphorylates the carboxy terminal domain (CTD) of pol-II (Marshall et al.,1996). We tested this in transgenic mouse fetal liver cells using *in situ* hybridization with probes that hybridize at different distances relative to the site of initiation of the β-globin primary transcript. *In situ* signals with probes that hybridize to intron 1, located in the first 300 bases of the primary β-globin transcript (Fig.1b) are still visible in 85% of the cells after 15 minutes of DRB treatment (Figure 2a,c), when compared to the non DRB treated control. Probes that hybridize to intron 2, 600-1200 bases 3' of the initiation site (Figure 2a) have completely disappeared after 7.5 minutes of DRB treatment (Figure 2b,c). The DRB results confirm earlier reports which suggested that in the presence of DRB, the process of transcription initiation is not disturbed but only elongation is affected resulting in short, prematurely aborted transcripts. Since initiation continues and the balance between γ and β is maintained we conclude that whichever mechanism (scanning, accessibility or looping) is responsible for the activation of the genes, it is not disturbed by the addition of DRB. The result also shows that after inhibition of transcription by DRB the time required to decay the existing β-globin intron 2 primary transcript signal via splicing, to levels beyond the limit of detection is in good agreement with the actinomycin-D results (Fig.1c,d). The same is found for the α-like and the γ genes (not shown).

### Kinetic analysis of single and double primary transcript signals

Primary transcript *in situ* hybridization with gene-specific intron probes for human γ- and β-globin in transgenic 11.5 day fetal liver cells containing a single copy of the complete human β-globin locus show single gene transcription signals in approximately 85% of the human globin loci (Wijgerde et al., 1995). A small percentage of loci (~15%, eg. fig. 3a, bottom right) contain signals for both γ- and β-globin genes *in cis* and it is these signals that make a distinction between the different mechanisms difficult. If double signals are due to simultaneous initiation of the γ- and β genes (on a random basis) according to the accessability and scanning models then release of the DRB block should result in the reappearance of double signals at the same rate as single signals. A lag in the reappearance of double signals would be indicative of single gene initiation. Double signals would result from the overlap between decaying primary transcripts (7.5 minutes, see above) from a recently active gene and the nascent transcription of an active gene as predicted from the alternating initiation of genes. Thus release from the DRB block in that case would predict an early reappearance of single gene signals (using intron 2 probes) followed by the delayed reappearance of double gene signals.

We treated transgenic 11.5 day homozygous fetal liver cells with DRB for 15 minutes to block elongation of globin primary transcripts (Fig.3 a, b). The cells were then released from the DRB block by washing with PBS and aliquots of cells were fixed onto slides at various intervals and prepared for *in situ* hybridization as previously described (Wijgerde et al., 1995). The cells were probed with intron 2-specific probes for human γ- and β-globin primary transcripts. No transcription signals are visible in the zero time point slide immediately after the wash (fig. 3b). Five minutes after washing out the DRB (figure 3c) primary transcript signals are again detectable in a high proportion of cells and continue to rise to the level observed before addition of DRB. Single (γ or β) and double (γ and β *in cis*) gene signals were counted and are presented in Figure 4. The results from three independent experiments show that the reappearance of double gene signals significantly lags behind the reappearance of single gene signals in the population. To determine whether there is an intrinsic bias against the appearance of double signals we counted the reappearance of two signals in trans (i.e. a signal on each chromosome) as an internal control. Their reappearance (fig.4b) closely approximates the single signal curves. We also calculated a theoretical curve that would be expected when the two loci are activated independently in trans (not shown). This curve coincides with the curve found for the reappearance of two signals in trans. We therefore conclude that the two allelic globin loci behave completely independently from each other in terms of transcription (showing a random reappearance of

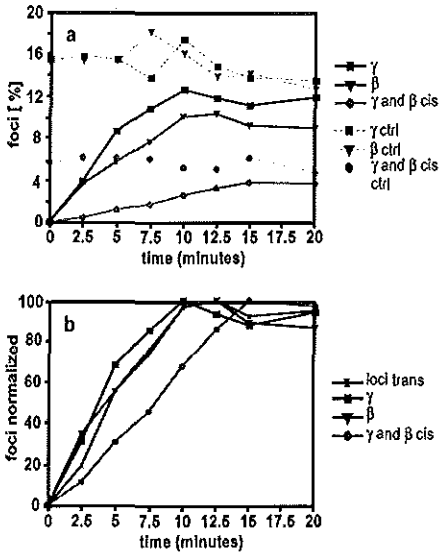



**Figure 4.** Kinetics of reappearance of single versus double gene signals. Panel a; γ- and ß-globin in situ signals (see fig 2) were scored after release of the DRB block to transcription elongation in 11.5 day fetal liver cells and plotted versus time. Solid lines DRB treated cells. Broken lines non DRB treated control. Note that the time of incubation only leads to a small decrease in the number of transcription signals. Greater than 1000 cells were counted for each time point. Panel b, The curves in panel a were normalized to their maximum values. The results show the early reappearance of foci having single γ- or ß-globin signals and a clear lag in reappearance of double (γ and ß in cis) signals. The experimental curve, loci in trans is derived by quantitation of the reappearance of cells with γ or ß signals on both chromosomes. It approximates the single signal curves and shows random appearance.

two signals without a lag) whereas the genes within a locus on one chromosome do not behave independently (double signal reappearance is non random with a lag).

### Faster alternation of initiation leads to a faster appearance of double signals

In the analysis described above the γ-genes are treated as one entity, however they are two independent genes, $^{G}\gamma$ and $^{A}\gamma$ which are expressed in embryonic red cells and early fetal liver erythroid cells of transgenic mice (Strouboulis 1994). Restriction digestion of sites in the γ promoter regions in isolated K562 nuclei indicated that both promoters were accessible to digestion *in cis* in approximately 50% of the loci (Bresnick et al., 1994). Co-accessibility of the promoters was inferred to be synonymous with simultaneous nuclease hypersensitivity of the two gene promoters. Although these links have not formally been established, it was suggested that this was an indication of co-initiation of transcription of the genes.

The two γ-genes are highly homologous in sequence throughout the coding, non-coding and intronic regions, making the use of primary transcript probes extremely difficult. We therefore took advantage of the fact that gene transcription proceeds well beyond the cleavage/polyadenylation site and terminates at a considerable distance downstream of the γ- genes in the 3' flanking region (Ashe et al., 1997) in which the two gene sequences diverge. Antisense oligo probes which recognize the regions 100-500 bp downstream of the polyadenylation sites were designed which specifically detect the two genes separately (not shown). Actinomycin-D experiments (fig. 5a) indicate that the RNA *in situ* signal in this region of the transcription unit has a shorter lifetime (3-4 minutes) than the intronic regions. This shorter lifetime is in principle an advantage as it allows more precise kinetic analysis. The primary transcript *in situs* for $^{G}\gamma$ and $^{A}\gamma$ on 11 day embryonic blood show that 15% and 3% are single $^{G}\gamma$ and $^{A}\gamma$ signals respectively, and 72% are double $^{G}\gamma$ and $^{A}\gamma$ signals *in cis* (figure 5). Approximately 10% of the loci show no γ signal which presumable indicates ε only transcription (Wijgerde et al;., 1995). The shorter lifetime of the *in situ* signals and higher percentage of double signals *in cis* suggests that these genes are possibly alternating faster than the γ-genes and the β gene (Wijgerde et al., 1995; Dillon et al., 1997). Thus the kinetic analysis of the reappearance of double $^{A}\gamma$ and$^{G}$ γ signals *in cis* should show little or no lag especially when the mechanics of DRB inhibition of elongation are taken into account. DRB inhibited complexes in vitro stall on the template and fall off after a few minutes thus allowing new polymerases to occupy the stall site (Marshall et al., 1992). Therefore if the $^{A}\gamma$ and$^{G}$ γ genes alternate more rapidly, stalled polymerases may be present on both genes simultaneously in a higher % of cells than when comparing γ and β due to a delay in dissociation from the template. This would have the net effect of further reducing the lag in reappearance of double signals *in cis*.

The kinetic analysis of transcription signal reappearance after DRB block and release shows that the double signals for $^{A}\gamma$ and $^{G}\gamma$ lag behind the single signals (Figure 6). Although not as pronounced as in the case of γ and β (Figure 4), the same result was obtained in three independent experiments and shows that the single signals peak earlier. On first sight this result maybe interpreted to show that the genes are co-initiated. However such a mechanism would not only fail to explain the lag in appearance, but also the fact that the single signals subsequently decrease as the double signals reach their maximum. The decrease shows that single signals are converted into double signals as would be expected in an alternating mechanism of transcription initiation, but not a co-initiation mechanism. This is also seen in the γ and β results, but to a much lower extent. Again this would be expected because there is a much lower percentage of double signals (15% in case of γβ vs 72% in case of $^{A}\gamma^{G}\gamma$) resulting in a much lower conversion from single to double signals. These results are similar to those observed in comparable experiments on the endogenous mouse β-like globin genes (T. Trimborn submitted) which appear to be controlled by a similar mechanism.

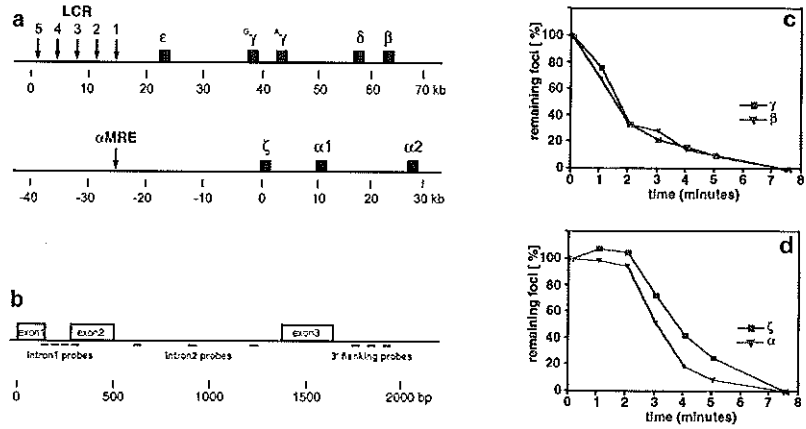



**Figure 1.** Schematic diagrams of Panel a, the human ß-globin locus construct used to make transgenic mice and the mouse locus and Panel b, intron/exon structure of a ß-like globin gene with localization of probes used for in situ hybridizations. Panel c shows the decay of the primary transcript signal after treatment with actD. The remaining intron2 signals of the γ- and ß-transcripts are shown as a function of time and have been normalized to 100% for the number of transcription signals at the start of the experiment (>95% of the alleles have a transcription signal) and a minimum of 600 cells were counted. Panel d shows the same as panel c for the ζ- and α-genes.

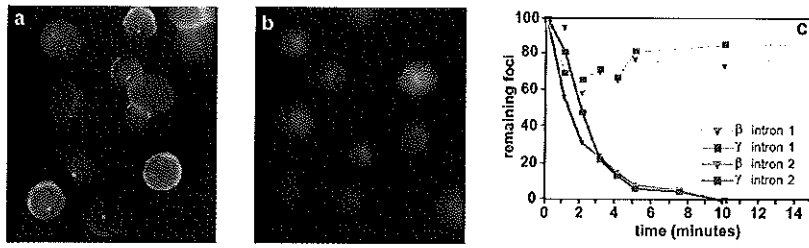



**Figure 2.** Panel a: In situ hybridization with intron 1 specific probes on heterozygous transgenic 11.5 day disrupted fetal liver cells treated for 15 minutes with 100 µM DRB. γ-globin signals are shown in red (Texas red), and ß-globin signals are green (FITC). Panel b: In situ hybridization with intron 2 specific probes on heterozygous transgenic 11.5 day disrupted fetal liver cells treated for 7 minutes with 100 µM DRB. γ-globin signals are red (Texas red), and ß-globin signals are green (FITC). Panel c: Decay of γ-intron 1 and intron 2 plotted as a function of time after the addition of DRB, γ-intron 1, broken red line; γ-intron 2, solid red line, ß-intron 1 broken green line, ß-intron 2 solid green line.

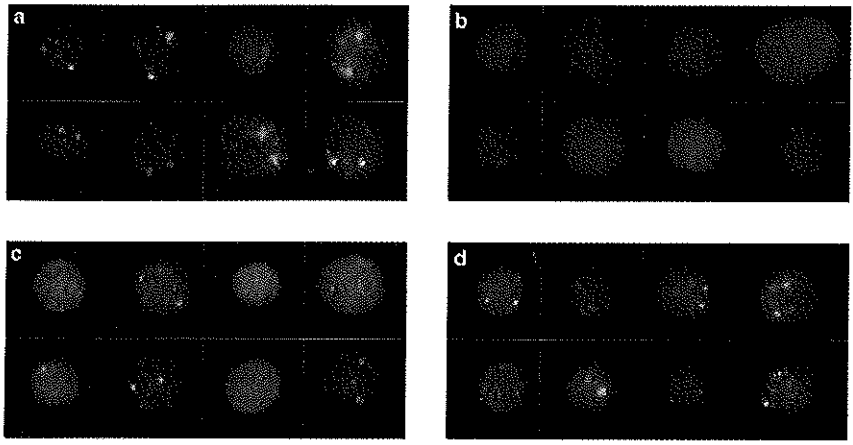



Figure 3 Reversible inhibition of transcription elongation on homozygous 11.5 day transgenic fetal liver cells. a-d, Primary transcript in situ hybridizations using γ-globin intron 2 probes, signals are shown in red (Texas red) and ß-globin intron 2 probes, green (FITC). a, γ- and ß-globin signals prior to DRB treatment. b, 15 minutes treatment with 100 µM DRB. c, 5 minutes after release of the transcriptional elongation block by washing out DRB. Note most cells show single γ- or ß-globin transcription signals in cis, while many have more than one signal in trans. d, 20 minutes after release of elongation block, the distribution of foci having single γ or ß and double signals (γ or ß in cis) are back to control levels. Representative cells are shown for each time point.

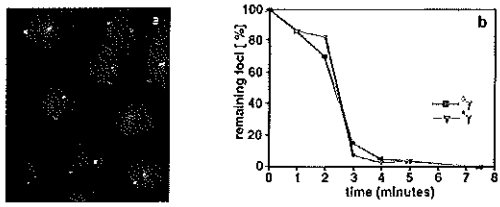



**Figure 5.** Panel a: Primary transcript in situ hybridization on 11.5 day embryonic blood cells with $^{A}\gamma$-globin and $^{G}\gamma$-globin 3' flanking region specific probes (see fig. 1b). $^{A}\gamma$-globin is shown in green (FITC), $^{G}\gamma$-globin is shown in red (Texas red). Panel b: Decay of the 11.5 day primary transcript signal of the γ globin genes using a 3' flanking probe (fig.1b) after the addition of actD. The remaining $^{A}\gamma$- signal is plotted in green, $^{G}\gamma$ in red, as a function of time.

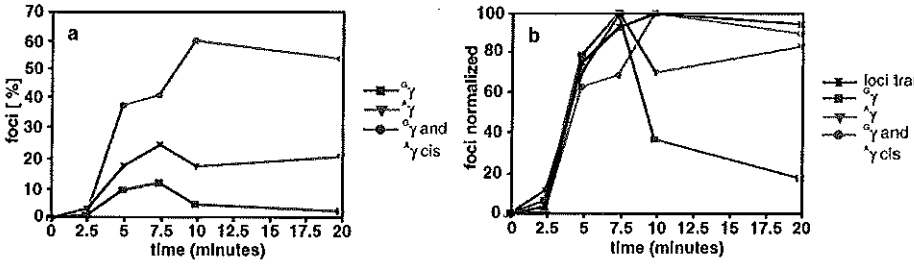



**Figure 6.** Kinetics of reappearance of single versus double $^{G}\gamma$- and $^{A}\gamma$-globin signals. a, In situ signals (similar to experiments shown in figure 3) were scored after release of the transcriptional elongation block in 11.5 day embryonic blood cells and plotted versus time. Greater than 1000 cells were counted for each time point. b, Curves were normalized to their maximum values for single ($^{G}\gamma$, or $^{A}\gamma$) and double ($^{G}\gamma$ and $^{A}\gamma$ in cis) signals. The curve $^{G}\gamma$ and $^{A}\gamma$ in trans, was derived as in figure 3 and follows a random curve.

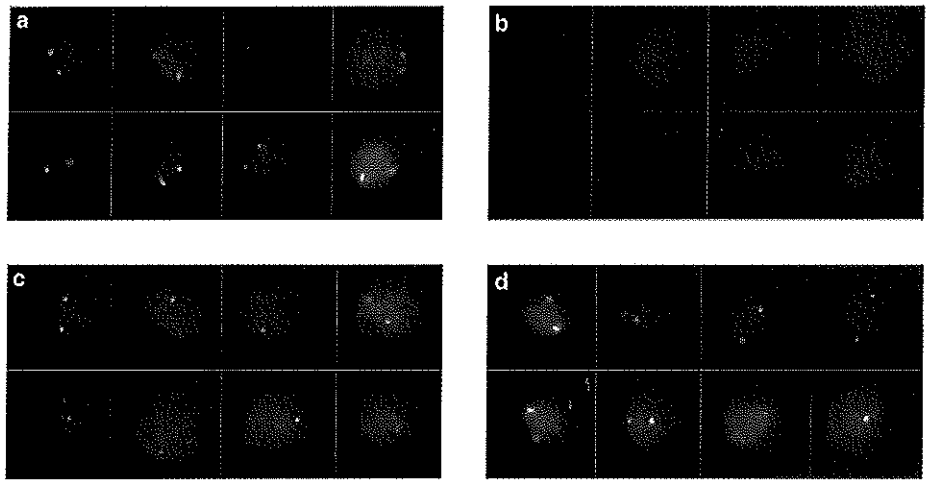



**Figure 7.** Reversible inhibition of transcription elongation on 10.5 day blood cells. a-d, Primary transcript in situ hybridizations using ζ-globin 3'intron 1, intron 2 probes, signals are shown in red (Texas red) and α-globin intron 2 probes in green (FITC). a, ζ- and α-globin signals prior to DRB treatment. b, 15 minutes treatment with 100 μM DRB. c, 5 minutes after release of the transcriptional elongation block by washing out DRB. Note that many cells show double ζ- and α-globin transcription signals in cis. d, 20 minutes after release of elongation block, the distribution of loci having single ζ or α and double signals (ζ or α in cis) are back to control levels. Representative cells are shown for each time point.

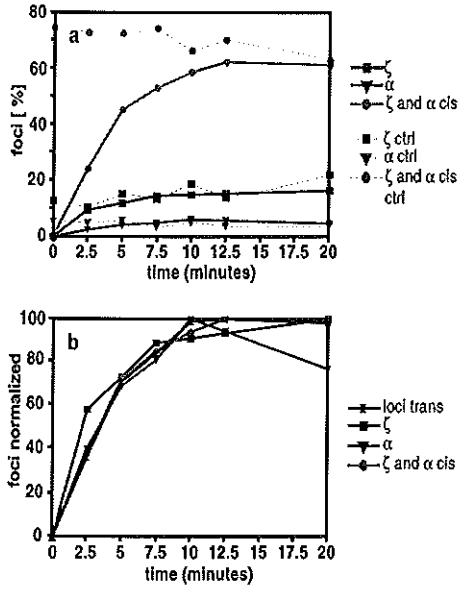


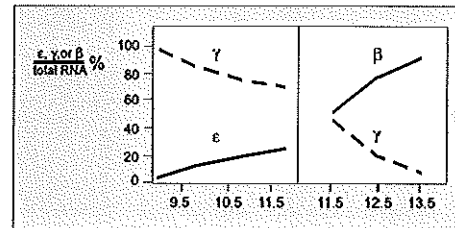

**Figure 8.** Kinetics of reappearance of single versus double ζ- and α-globin signals. Panel a; In situ signals (similar to experiments shown in figure 3) were scored after release of the transcriptional elongation block in 10.5 day embryonic blood cells and plotted versus time. Greater than 1000 cells were counted for each time point. Panel b, Curves were normalized to their maximum values for single (ζ or α) and double (ζ and α in cis) signals. The random curve (loci in trans) was derived as in figure 3.

### The amount of $^{G}\gamma/^{A}\gamma$ double signals changes during development

The percentage of $^{G}\gamma$/double signals also provides a further means of distinguishing the different models of long range transcriptional activation. The ratio of $^{G}\gamma$ to $^{A}\gamma$ transcription changes very little during development (39% $^{G}\gamma$ of total γ on day 9.5 to 26% on day 11.5, Table 1) in the transgenic mice. Thus if co-initiation of transcription occurs, it would predict that the percentage of $^{G}\gamma/^{A}\gamma$ double signals in the γ expressing cells would change very little. In contrast if alternate transcription occurs, it would predict that the ratio should decrease in the γ expressing cells from early embryo, when γ transcription is accompanied by a low level of ε transcription, to the fetal liver when γ transcription is accompanied by a high level of β transcription (Strouboulis et al., 1992). When the percentage of $^{G}\gamma/^{A}\gamma$ double signals is measured in the γ-expressing cells during development, it changes from 83% at day 9.5 (Table 1) to 56% at day 11.5 in embryonic blood (Fig. 5a, Table 1) to 29% at day 11.5 in the fetal liver (Table 1). Thus when γ- gene expression is accompanied by the highly expressed β-gene in the locus, the percentage of $^{G}\gamma/^{A}\gamma$ double signals decreases as predicted by alternate single gene transcription.

### The transcription initiation of α-like globin genes

The analysis above does not exclude that the treatment with DRB would somehow artifactually cause a delay in the reappearance of double signals *in cis* for some unknown reason, even though this is clearly not the case for the reappearance of double signals in trans (fig. 4b an 6b). We therefore examined the primary transcription of the α genes. There is indirect evidence that the α-globin genes may be regulated differently than the β-like genes. Instead of the five hypersensitive regions present in the β-globin LCR, only a single hypersensitive site has been identified 40kb or 26kb upstream of the human and mouse genes respectively (fig.7, Gourdon et al., 1995). When this site is present as part of the human α-locus in transgenic mice, the expression of the α-genes is suppressed as development procedes (Gourdon et al., 1993; Sharpe et al., 1993). In contrast when the α-globin genes are placed under the control of the β-globin LCR no such suppression is observed (Hanscombe et al., 1989; Ryan et al, 1989). This indicates that the α-locus may normally be present in a constitutively active region of the chromatin (Craddock et al., 1995) or that there are as yet undefined additional regulatory sequences (Gourdon et al., 1994, 1995), opening up the possibility that the genes may not be regulated by alternating initiation of transcription. Primary transcript *in situ* analysis shows a high percentage of double ζ/α signals *in cis* (80%) versus single ζ or α signals (fig.7a), while the lifetime of the signal is similar to that observed for the β-signals (fig. 1d) in the presence of actinomycin-D. This could mean that the genes are frequently co-initiated or that alternating initiation




| | % in situ primary γ RNA signals | | | % total γ RNA | |
|---|---|---|---|---|---|
| | Gγ | Gγ/Aγ | Aγ | Gγ | Aγ |
| day 9.5 blood | 5 | 83 | 12 | 39 | 61 |
| day 10.5 blood | 8 | 78 | 14 | 31 | 69 |
| day 11.5 blood | 7 | 56 | 37 | 27 | 73 |
| day 11.5 liver | 23 | 29 | 48 | 26 | 74 |

**Table 1.** The table shows the % of cells expressing γ with a $^{G}\gamma$ or $^{A}\gamma$ single or a $^{G}\gamma/^{A}\gamma$ double primary transcript signal by in situ hybridization (bottom left) on different days of development. The bottom right shows the % of $^{G}\gamma$ and $^{A}\gamma$ of total steady state RNA in the pool of cells as determined by primer extension (data not shown). The top panel illustrates the change in the levels of ε, γ and ß RNA as a % of total human globin RNA in the transgenic mice (Strouboulis, 1994). Please note that the total RNA levels (accumulated mRNA) lag behind the in situ transcription levels (primary transcription).

takes place rapidly. If the latter takes place well within the lifetime of the signal a large proportion of double signals *in cis* would result. We then measured the reappearance of single and double signals after the addition and removal of DRB (fig. 4b, c and d). Plotting the reappearance of the signals as a function of time (fig. 8) shows that the double signals appear at the same rate as the single signals and that there is no measurable conversion of single signals into double signals. On the basis of these data it is tempting to suggest that the α-like genes are co-initiated, but rapidly alternating initiation of transcription would give a similar result and hence can not be excluded. It is difficult to fully explain the data obtained with the α-locus in mechanistic terms, because not all the regulatory regions in this locus have been identified (Gourdon et al., 1995) leaving the possibility that the different genes could be activated simultaneously by different regulators.
Importantly the result shows that the treatment with DRB does not artifactually result in a lag in the appearance of double signals *in cis*.

## DISCUSSION

The results obtained previously with primary transcript analysis suggested that the the human β-globin genes are regulated via a dynamic process of alternating initiation of transcription of the different genes. That proposal appeared to contradict existing data. Firstly it had been known for many years that γ- and β-mRNA and proteins are found in the same cell during the switchover from γ to β, even in heterozygotes and this appeared to be confirmed by single cell PCR data (Furukawa et al., 1994). Secondly restriction enzyme analysis of the globin locus in nuclei showed that both γ-globin promoter regions could be cleaved releasing the intervening 5kb fragment. These data were interpreted to mean that the γ- and β-globin genes within a single locus were or could be transcribed simultaneously (Martin et al., 1996; Bresnick et al., 1994; Furukawa et al., 1994). On the basis of the presence of (a minority of) loci with double primary transcription signals *in cis* the *in situ* primary transcription data have been interpreted to suggest that co-initiation takes place (Martin et al., 1996).

It is therefore important to distinguish whether alternating single or multiple co-initiation takes place, because they have direct implications for our understanding of the process of transcriptional initiation *in vivo* involving distant regulatory sequences. Three models have been proposed to explain transcriptional regulation by distant regulatory sequences. Firstly an accessibility model where after the activation of the chromatin, the genes bind transcription factors and are activated in a stochastic fashion (Groudine and Weintraub, 1982; Martin et al., 1996). Secondly a scanning model where the regulatory sequences act as a nucleation site for (part of) the transcription machinery, which subsequently scans the DNA for transcriptionally competent genes (Herendeen et al., 1992; Tuan et al., 1992). Thirdly a looping model which postulates that the distant regulatory sequences interact directly with the gene to initiate transcription (see Ptashne, 1988 and Wijgerde et al., 1995 and ref therein). Unfortunately it is at present not possible to directly visualize the transcriptional process *in vivo* and distinguish between these mechanisms. Hence it has only been possible to design experiments which are capable of verifying predictions made from each of the models. In case of the β-globin gene locus a number of *in vivo* observations have to be accounted for in each of the proposed mechanisms.

For the purpose of this discussion the most important points about the regulation of the β-globin locus are that a single regulatory region (LCR) is required by all the genes in the locus (reviewed in Grosveld et al., 1993), that the genes compete with each other (Giglioni et al., 1984; Enver et al., 1990; Behringer et al., 1990; Hanscombe et al., 1991; Dillon et al., 1997) and that a gene closer to the LCR has a competitive advantage over a more distal gene (Hanscombe et al., 1991; Peterson and Stamatoyannopoulos, 1993; Dillon et al., 1997).

The accessibility model does not allow for competition and hence an additional mechanism, interference with a downstream gene, was postulated to give proximal genes a competitive advantage over distal genes (Martin et al., 1996). The model predicts that two or more genes would initiate transcription at the same time in a substantial proportion of the cells. The double signals in our primary transcript analysis were therefore taken as evidence that co-initiating loci were present in a substantial number of cells. However the data presented here show that the double signals are not generated as predicted by a co-initiation mechanism, but represent a transition between alternating transcriptional periods of single genes. This argues strongly against the accessibility model. Moreover this essentially stochastic model does not explain why the number of double signals is higher when two γ-genes *in cis* (fig. 5 and 6) are compared to a γ and β signal *in cis* (fig. 3 and 4), while their expression is lower than the β-gene (Table 1). In addition it fails to explain why the percentage of $^{G}\gamma/^{A}\gamma$ double signals decreases during development when the other genes in the locus change from low expression (ε) to high expression (β) (Table 1). It also predicts that there should be a substantial proportion of loci without a primary transcript signal. Such a proportion should be approximately 15% on a stochastic basis when 85% and 15% of the loci show single or double signals respectively. However 3% is found (data not shown and Wijgerde et al., 1995). This model also predicts the opposite result to what is observed when the distance between the genes is changed (Dillon et al., 1997). Thus it very difficult to see how this model could explain the regulation of this locus *in vivo*. If this were the mechanism a substantial number of complex additional assumptions would have to be made.

The scanning model does allow for a competitive advantage of proximal over distal genes, because a scanning process initiated at the LCR would encounter a proximal gene first. However it would also predict that co-initiation of transcription would

frequently take place, which the data presented here show is not the case. It also does not explain why the percentage of $^{G}\gamma/^{A}\gamma$ double spots decreases during development in the γ expressing cells. The model predicts that even when the expression of γ decreases the percentage of $^{G}\gamma/^{A}\gamma$ double spots should stay the same when the ratio of $^{G}\gamma/^{A}\gamma$ expression remains almost the same (see above, Table 1). The scanning model also fails to explain why the expression of the γ genes is increased when the expression of the β gene is decreased due to the absence of a EKLF, a β-gene specific transcription factor (Wijgerde et al., 1996). It would predict that the level of γ gene expression would stay the same. Lastly it is known that a β-globin gene situated close to the LCR has the same output as a β-globin gene at a much larger distance (Dillon et al., 1987 and references therein). A scanning model would therefore predict that gene order but not distance would be an important parameter. However the results obtained by varying distances show that distance is important (Hanscombe et al., 1991; Dillon et al., 1997; Peterson and Stamatoyannopoulos, 1993). When all these observations are taken together they argue very strongly against a scanning mechanism operating *in vivo*, unless extra complex parameters are introduced. Nevetheless it is important to note that this does not mean that a scanning process could not be involved in the first activation of the locus. In the experiments to date only the process of multiple rounds of transcription (reinitiation) after initial activation of the locus has been examined.

In contrast to the two models discussed above, a dynamic interaction between the LCR and the genes via looping does explain all of the basic properties. The results presented here show that RNAs from multiple globin genes do not co-initiate *in cis*, which therefore establishes a central aspect of the regulation of the globin locus. Co-expression of more than one globin gene *in cis* would be achieved by alternating transcriptional periods of single genes. It explains competition, because it predicts that the time taken up by LCR driven transcription of one gene takes time away from another gene. In this model the frequency of LCR/gene interactions would be dependent on distance with an advantage of proximal genes over distal genes (see Dillon et al., 1997). The dynamic interaction model by looping also makes clear why the percentage of $^{G}\gamma/^{A}\gamma$ double spots decreases during development (Table 1). Late in development, the LCR is contacting the competing β gene most of the time in the γ expressing cells, whereas the opposite is true for the ε gene early in development (Strouboulis et al., 1992). Thus the chance that the LCR interacts with one γ gene shortly after interacting with the other γ gene has decreased and hence the chance that one γ signal has decayed before the synthesis of the other γ is much higher late in development. The double spots which are the result of the overlap between decaying and newly synthesized γ-signals would therefore decrease. For very similar reasons it predicts that the number of γ double signals (Fig. 5 and 6) is higher than the number of γ/β double signals (Fig 3 and 4). It suggests the LCR can rapidly move between γ genes early in development (hardly disturbed by the competing ε gene) resulting in a lot of signal overlap. Late in development the LCR interacts with β for much longer times allowing γ-primary transcripts to decay and thus decrease the amount of overlap signals. Finally the dynamic interaction model also makes clear why a decrease in the amount of EKLF results in an increase in γ expression (Wijgerde et al., 1996). The LCR/β gene complex would be destabilized by the decrease or absence of this β-gene specific factor and hence the frequency of competing γ gene interaction would increase, resulting in an increase in γ-gene initiation of transcription.

Looping and direct contact between regulatory regions is therefore the most simple mechanistic explanation for the observed results and is supported by previous experiments with other systems (Mueller-Storm et al., 1989; Bickel and Pirotta, 1990; Dunaway and Dröge, 1989). It implies that direct chromatin interactions between the LCR and a single gene are required for initiation of transcription. Continued loading of polymerases or reinitiation of that gene would require continuous LCR contact (Wijgerde et al., 1995; Milot et al., 1996; Dillon et al., 1997). In the context of this mechanism the data suggests that chromatin *in vivo* is highly dynamic or diffusible, allowing the LCR/gene complex to change rapidly to bring about co-expression of multiple genes.

## Methods

### Reversible inhibition of transcription elongation with DRB.

Homozygous transgenic mice containing a single integrated copy of the complete human β-globin locus were bred to obtain embryos. Peripheral blood and fetal livers were dissected out in PBS. 10.5 or 11.5 day embryonic blood and 11.5 or 12.5 day fetal livers were disrupted into PBS by repeated pipetting as previously described (Wijgerde et al., 1995). DRB (Sigma) was added to the cell suspensions to a final concentration of 100 μM or (Actinomycin-D at 5 μg/ml) and incubated at 37°C. for 15 minutes. Five volumes of ice cold PBS were added and the cells were immediately pelleted by centrifugation for 2 minutes at 1500 rpm in an eppendorf centrifuge. Cells were washed two more times with 1.5 ml ice cold PBS and resuspended in 250 ml of PBS at 25°C. Aliquots were taken at the designated intervals and fixed onto poly-L-lysine coated slides (Sigma) for *in situ* hybridization.

### *In situ* hybridization analysis

The following probes were used for the *in situ* hybridization analysis:

Human β intron 1 probes:
5'-CTGTCTCCACATGCCCAGTTTCTATGGTCTCCTTAA ACCTG TCTTGTAA-3'; 5'GGGTGGGAAAATAGACCAAA GGCAGAGAGA GTCAGTGCCTATCAGAAAC-3'; 5'-AGG GCAGTAACGGCAGACT CTCCTCAGGAGTCAGGT-3'; 5'-ATAACAGCATCAGGAGGGACA GATCCCCAAAGGA CTCA-3'

Human β intron 2 probes:
5'-TTCCACACTGATGCAATCATTCGTCGTTTCCCATTC TAAACTGTACCCT-3'; 5'-CTGATTTGGTCAATATGTGT ACACAATTAAAACATTACACTTTAACCCA-3'; 5'-GGTAG

CTGGATTGTAGCTGCTATAGCAATATGAAACCTCTTAC ATCAGT-3'
Human γ intron1 probes:
5'-AGGCACAGGGTCCTTCCTTCCCTCCCTTGTCCTGG TCAC-3'; 5'-TGACAAGAACAGTTTGACAGTCAGAAGGT GCCACAAATAATGAGAAGCGA-3'; 5'-AGGCTTGTGATA GTAGCCTTGTCCTCCTCCTCTGTGAAATGACCCA-3'; 5'-AGAGCCTACCTTCCCAGGGTTTCTCCTCCAGCATCTTC CACATT-3'
Human γ intron2 probes:
5'-GCAGTTTCTTCACTCCCAACCCCAGATCTTCAAACA GCTCACACCCGC-3'; 5'-CCTTCTGCCTGCATCTTTTTAA CG ACCAACTTGTCCTGCCTCCAGAAG-3'; 5'-ACAGAGC TGACTTTCAAAATCTACCCAGCCCAAATGTTTCAATTG TCC-3'
Human $^{G}$γ 3' flanking probes:
5'-ACGTAAACAAAAAAGTGTGGAGTGGCACATGACAC AAACACACATAG-3'; 5'-GCAGACGCTCCCATGTATAAGT TCTTTATTTGCCTAGTTCTTTTATTT-3'; 5'-GTTGAGCCC CTTCCTCGCTGCTGGCTGCAATCCAGGGAAGGGGGTT C-3'
Human $^{A}$γ 3' flanking probes:
5'-TCATATAAAAATAAATGAGGAGCATGCACACACCAC AAACACAAACAGGC-3'; 5'-AGAACTCCCGTGTACAAGT GTCTTTACTGCTTTTA-3'; 5'-TTCATTAAGAACCATCCTT GCTACTCAGCTGCAATCAATCCAGCCCCCA-3'; 5'-ATTT CACTTTCTTAGGCATCCACAAGGGCTGTGAAAAGCTAA GTGCCAT-3'
Mouse α probes:
5'-CACAGAAAAGCATAGTTAGAAGCGCCCACTGAGCG AGTGCCAGGTCC-3'; 5'-AGCCCTTCCTAGGGGCCCAG ATGCCGCCTGCCAGGTCCC-3'; 5'-GCTCCCCTTCCTGG GACCACTATGTCCCTGCCTTGGGCACGAGGACCC-3'
Mouse ζ probes:
5'-CCTTCTCAGTGGCTTCTCCTCACAACTGCTCTTTGT CACTTCTGTCTC-3'; 5'-ATGGAAGACTCTGGTGAGCTCT GGAATGCCAGCCCACCTCCTTTAGTA-3'; 5'-ACAACCC CAAGAGTGATGTTACTATTGCTGTTGCACAAGGGTCTA CA-3'; 5'-AAGGGGATTTGATGCCTCCAGCCCCAATGGC ACCCATGCCTGCGCTCG-3'

Cells were fixed onto a poly-L-lysine coated slide in 4% formaldehyde/5% acetic acid for 18 min. at room temp. The cells were subsequently washed 3 times 5 min in PBS and stored in 70% ethanol at -20°C. The slides were pretreated for hybridization by a 0.01% pepsin digestion (5 min., 37°C) in 0.01M HCl, followed by a short wash in water and a 5 min. fixation in 3.7% formaldehyde at roomtemp. The slides were washed in PBS, dehydrated in 70%, 90% and 100% ethanol steps and air dried. The hybridization mixture was applied (12 µl per 24x24mm coverslip) and incubated at 37°C in a moisturized chamber for 12hr. The hybridization mixture contained 1ng/µl of each of three or four oligonucleotides (50nt) containing either a digoxygenin or a biotin side chain in the middle and on the 5' and the 3'end of the oligonucleotide (Eurogentec, Belgium) in 25% formamide, 2x SSC, salmon sperm DNA (200ng/µliter), 5x Denhardts, 1mM EDTA and 50mM sodium phosphate pH 7.0. The coverslip was removed by dipping in 2x SSC and the cells were washed three times in 2x SSC 37°C, followed by a 5 min wash in 0.1M Tris, 0.15M NaCl, 0.05% Tween 20. Antibody detection of the labels was essentially as described by Dirks et al. (1993), with three or four amplification steps. Mounting was in DAPI/DABCO:Vectashield (1:1) in glycerol (90%) and stored at 4°C in the dark. Fluorescence was detected by epifluorenscence/CCD.
In all cases more than 500 cells were counted per data point. The count was confirmed blindly by an independent person using a dictaphone.

**References**


Ashe, H., Monks, J., Wijgerde, M., Fraser, P., Proudfoot, N. Intergenic transcription and transinduction of the human β-globin locus. *Genes Dev.* in press (1997).

Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L. and Townes, T.M. Human gamma- to beta-globin gene switching in transgenic mice. *Genes Dev.* 4(3), 380-9 (1990).

Bickel, S. & Pirotta, V. Self association of the Drosophila zeste protein is responsable for transvection effects. *EMBO J* **9**, 2929-67 (1990).

Bresnick, E.H., Felsenfeld, G. Dual promoter activation by the human β-globin locus control region. *Proc. Natl. Acad. Sci. U S A.* **91**, 1314-1317 (1994).

Chodosh, L.A., Fire, A., Samuels, M., Sharp P.A. 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *J. Biol. Chem.* **264**, 2250-2257 (1989).

Craddock, C.F., Vyas, P., Sharpe, J.A., Ayyub, H., Wood, W.G., Higgs, D.R. Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments. *EMBO J* **14**(8), 1718-26 (1995).

Dillon, N. and Grosveld, F. The human γ globin gene is silenced independently from the other genes in the β- globin locus. *Nature* **350**, 252-254 (1991).

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. and Grosveld, F. The effect of distance on long range chromatin interactions. *Molec. Cell* **1**, 131-138 (1997).

Dirks, R., van de Rijke, F., Fujishita, S., van der Ploeg, M. and Raap, A. Methodoligies for specific intron and exon RNA localisation in cultured cells by haptenized cells and fluorochromized probes. *J. of Cell Sci.* **104**, 1187-1197 (1993).

Dunaway, M. and Dröge, P. Transactivation of the Xenopus rRNA gene promoter by its enhancer. *Nature* **341**, 657-659 (1989).

Enver, T., Raich, N., Ebens, A.J. et al. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* **344**, 309-13 (1990).

Foley, K.P., Engel, J.D. Individual stage selector element mutations lead to reciprocal changes in beta- vs. epsilon-globin gene transcription: genetic confirmation of promoter competition during globin gene switching. *Genes Dev.* **6**,730-744 (1992).

Fraser, N.W., Sehgal, P.B., Darnell, J.E. DRB-

induced premature termination of late adenovirus transcription. *Nature* **272**, 590-593 (1978).

Furukawa, T., Zitnik, G., Leppig, K., Papayannopoulou, T. and Stamatoyannopoulos, G. Co-expression of gamma and beta globin mRNA in cells containing a single human beta globin locus: results from studies using single-cell reverse transcription polymerase chain reaction. *Blood* **83**(5), 1412-9 (1994).

Giglioni, B., Casini, C., Mantovani, R. et al. A molecular study of a family with Greek hereditary Persistance of Haemoglobin and β-thalassaemia. *EMBO J.* **11**, 2641-2645 (1984).

Grosveld, F., Dillon and Higgs, D. The Regulation of Human Globin Gene Expression. In: Balliere's Clinical Haematology: The Haemoglobinopathies. Vol. 6/No. 1, pp. 31-66 (1993).

Gourdon, G., Sharpe, J.A., Higgs, D.R. and Wood, W.G. The mouse alpha-globin locus regulatory elements. *Blood* **86**(2), 766-75 (1995).

Gourdon, G., Sharpe, J.A., Wells, D., Wood, W.G., Higgs, D.R. Analysis of a 70 kb segment of DNA containing the human zeta and alpha-globin genes linked to their regulatory element (HS-40) in transgenic mice. *Nucl Acids Res* **22**(20), 4139-47 (1994).

Groudine, M., Weintraub, H. Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. *Cell* **30**, 131-139 (1982).

Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L., Greaves, D. R. and Grosveld, F. High-level, erythroid specific, expression of the human α-globin gene in transgenic mice and the production of human haemoglobin in murine erythrocytes. *Genes Dev*, **3**, 1572-1581 (1989).

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

Herendeen, D.R., Kassavetis, G.A., Geiduschek E.P. A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science* **256**,1298-1303, (1992).

Marshall, N.F., Price, D.H. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol.Cell.Biol.* **12**, 2078-2090 (1992).

Marshall, N.F., Peng,J., Xie, Z., Price, D.H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* **271**, 27176-27183 (1996).

Martin, D.I., Fiering, S., Groudine, M. Regulation of beta-globin gene expression: straightening out the locus. *Curr. Opin. Genet. Dev.* **6**, 488-495 (1996).

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

Mueller-Storm, H.P., Sogo, J.M., Schaffner, W. An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell* **58**, 767-777 (1989).

Peterson, K., Clegg, C., Huxley, C. et al., Transgenic mice containing a 248kb YAC carrying the human β-globin locus display proper developmental control of human globin genes. *Proc. Natl. Acad. Sci.* **90**, 7593-7597 (1993).

Peterson, K. and Stamatoyanopoulos, G. Role of order in developmental control of human gamma and beta globin gene expression. *Mol Cell Biol* **13**, 4836-43 (1993).

Ptashne, M. How eukaryotic transcriptional activators work. *Nature* **335**, 683-85 (1988).

Raich, N, Enver, T., Nakamoto, B. et al., Autonomous developmental control of human embryonic switching in transgenic mice. *Science* **250**, 1147-1149.

Sharpe, J.A., Wells, D.J., Whitelaw, E., Vyas, P., Higgs, D.R., Wood, W.G. Analysis of the human alpha-globin gene cluster in transgenic mice. *Proc Natl Acad Sci* **90**(23):11262-6 (1993).

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

Tuan, D., Kong,S., Hu, K. Transcription of the hypersensitive site HS:2 enhancer in erythroid cells. *Proc.Natl. Acad. Sci. USA.* **89**, 11219-11223. (1992).

Wijgerde, M., Grosveld, F., Fraser, P. Transcription complex stability and chromatin dynamics. *Nature* **377**, 209-213 (1995).

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

**Acknowledgements**: We wish to thank Drs. M. Antoniou and T. McMorrow for helpful discussions and L. Braam for animal care. J.G. was supported by the Jeantet Foundation and E.M. by the MRC Canada. This work was supported by the NWO Netherlands.

# Chapter 6

General discussion

## 6. General discussion

Globin gene switching is defined as the sequential activation of the different genes according to a precisely ordered program during development. Without the LCR there is no activation of the locus and no transcription of the β-globin genes, emphasising its importance in the regulation of these genes (Grosveld et al., 1987 and 1993). In transgenic studies, the LCR provides position independent copy-number dependent expression at levels equivalent to the endogenous globin genes. Only a complete LCR opens the locus into a fully active chromatin configuration in all erythroid precursor cells (Milot et al., 1996). The LCR has been postulated to achieve these functions through a dominant positive effect that overcomes negative position effects rather than through an insulating effect. Because the LCR is important for the activation of all of the β-globin genes, its localisation results in an intrinsic structural polarity, which has a profound effect on the developmental expression of the genes. Many of the *trans* and *cis* components involved in the regulation of the human β-globin genes have been identified. However, the molecular details of how the LCR influences gene activity over distance are still a topic of debate.

The prevailing competition model suggests that the LCR activates the different genes through looping out of intervening DNA sequences establishing direct LCR-gene interactions (Dillon et al., 1997; this thesis). Looping models that bring distally located sequences into direct contact with promoters have been widely favoured and there are strong indications that enhancers work through such interactions (Ptashne et al., 1988; Muller et al., 1989; Bickel and Pirotta, 1990; Muller and Schaffner, 1990). The accessibility model (Groudine and Weintraub, 1982; Martin et al., 1996) and the tracking model (Tuan et al., 1992; Herendeen et al., 1992) have been proposed as alternative mechanisms of LCR activation and developmental regulation of the β-globin genes. In contrast to these models, the looping model can most easily explain all the basic properties of gene activation in the human β-globin locus without additional assumptions.

The data described in each chapter has been considered at length in the discussion sections at the end of each relevant chapter. Because no additional data has been obtained since, the discussion in those chapters represents our current view and working model. I will therefore only recapitulate the results obtained in this thesis and discuss how the results apply to the known transcriptional activation mechanisms. Thereafter I will discuss some aspects of the RNA fluorescent *in situ* hybridisation (FISH) technique on which a large part of the research presented in the thesis is based. Since it is a technique in development, I will illuminate some aspects concerning its liability and some of its current and future applications. In addition the results obtained in Chapters 3 and 4 concerning the role of EKLF in adult β-globin gene expression need to be clarified to some extent, since they appear contradictory to some extent.

*The LCR holocomplex or looping model*

The ordered arrangement of the β-globin genes and the involvement of the LCR in activating all of the genes accommodate the locus with an intrinsic structural polarity, which became evident by the propensity of the LCR to favour the transcriptional activation of a more proximally located gene (Enver et al., 1990; Hanscombe et al., 1991; Dillon et al., 1997). Varying the relative distance of two competing genes from the LCR results in changes in the level of gene expression and equivalent changes in the number of transcriptional active genes in a way that was predicted for the kinetics of interactions through free diffusion of two tethered elements (Dillon et al., 1997). These observations support a stochastic activation mechanism of direct LCR gene interactions. In addition, the nearly balanced competition that was observed implies that the LCR is shared among the active genes suggesting that it functions as a single transcriptional activation entity (Dillon et al., 1997). This is in agreement

with previous observations which suggested that individual HS elements function in an additive manner to provide full LCR activity (Fraser et al., 1990 and 1993) and that the LCR HS2 and HS3 contain separable enhancer and chromatin opening activities respectively (Ellis et al., 1996). It was therefore postulated that the different LCR elements interact to form a single larger complex or 'holocomplex', which interacts directly with the individual globin genes. How the different LCR elements are structurally organised in such a holocomplex is not known.

The best way to validate this LCR-'holocomplex' or looping model is through direct visualisation of contacts between two regulatory sequences located at considerable distances from each other, e.g. between the LCR and the proximal elements of the globin genes. However this has yet to be achieved *in vivo*. Only indirect approaches have been used to make inferences of the transcriptional activation mechanism operating in the β-globin locus. A pivotal question in understanding the transcriptional mechanistics in the β-locus is whether only a single or multiple genes can be transcribed simultaneously. Co-detection of multiple human β-globin mRNA species within single MEL/human fetal erythroid cell hybrids containing a single human β-globin locus and embryonic primitive erythroid cells have been interpreted to indicate a co-transcription mechanism (Furakawa et al., 1994 and 1995). Co-localisation of multiple globin proteins has also been visualised in single erythroid cells using fluorescence immunohistochemical detection methods (Fraser et al., 1993; Strouboulis, 1994 thesis). However, both globin mRNAs and proteins are extremely stable in erythroid cells which makes these observations merely imply that a single erythroid cell is not predetermined to express either one gene singularly during its life span but that gene switching can take place within a single erythroid cell (Fraser et al., 1993). Dual promoter activation through sharing of LCR activity has been proposed based on the finding that restriction endonuclease ApaI digestion leads to a high yield of a fragment spanning both γ-gene promoters in the human erythroleukemia cell line K562 but not in Hela cells (Bresnick and Felsenfeld, 1994). Similar chromatin characteristics were shown for the human γ-genes in primitive erythroid cells of single human β-globin YAC transgenic mice (Furakawa et al., 1995). However we have shown that DNaseI hypersensitivity at the β-promoter region and endonuclease sensitivity to ApaLI of the adult β-globin gene are not always coupled to active gene transcription (Milot et al., 1996). This implies that a high level of sensitivity merely reflects the potential of a gene to become active rather than gene activity itself. In order to distinguish between single versus multiple gene transcription in the human β-locus, one has to look at the single cell level at a relatively short-lived event such as the transcription process of a gene.

We have analysed the dynamics of globin gene switching via an indirect approach using RNA FISH to visualise primary transcripts on a single cell basis in erythroid cells of single copy human β-globin locus transgenic mice (Strouboulis et al., 1992; Wijgerde et al., 1995; Chapter 2). Probe sequences were directed against intron regions of the different globin genes, which are known to have a half life of 3-4 minutes *in vivo* (Chapter 2). Transcriptional inhibition using actinomycin-D revealed that all intron signals disappear to below detectable levels after 7.5 minutes, suggesting that fluorescent foci only appear at the site of an active or very recently transcribed gene (Chapter 5).

The results in embryonic and fetal erythroid cells of homozygous transgenic mice show that most of the transcription signals are in fact single gene transcription signals and only a minority of cells has double gene signals. The great variety of transcriptional cell types detected suggested a dynamic, continuously changing system of gene activation. The presence of cells with accumulated levels of cytoplasmic β-mRNA having only primary transcript signals of the developmentally earlier γ-genes, indicated a dynamic gene switching mechanism in which the genes are alternately transcribed. In the looping model this would

suggest that the LCR alternate in the interaction between these competing genes in a flip-flop type mechanism. The minority of globin loci displaying double gene signals in *cis* were interpreted to represent a transitory switching event in which the old decaying signal as well as the new actively transcribed gene are simultaneously detected. However, the double gene signals were interpreted by others as evidence for co-transcription accommodating the accessibility model of initiation (Martin et al., 1996). We therefore performed additional analysis of globin gene transcription to resolve the kinetics of the double gene signals by using RNA FISH analysis of globin primary transcripts in conjunction with transcriptional inhibition and release experiments. DRB was used to reversibly inhibit the elongation of RNA polymerase II in fetal liver cells of single copy human β-globin locus transgenic mice, initiation was shown to be unaffected (Chapter 5). The data shows that upon release of inhibition of the human γ- and β-genes the number of single gene signals accumulate faster and peak earlier when compared to the double gene signals. This implies that the double γ/β-gene signals represent a transition state between alternate single gene initiation (Chapter 5). Similar analysis has been performed for the human γ- and the murine α-genes. However, the results were less decisive and further scrutiny will be necessary to allow discrimination between alternate and co-initiation for these genes.

Interestingly, these experiments have shown that during periods of gene switching the number of transcriptionally active genes corresponds with the level of gene expression (mRNA detection). This suggests that gene transcription in the β-globin locus is an all or nothing event in which a gene is either fully active or silent. For the looping model this implies that the level of gene expression within a single cell is determined by the frequency and duration of LCR-gene interactions, rather than through modulation of the density of polymerases or changes in the rate of transcription. In addition, a continuous LCR-gene interaction is required for continual gene transcription and altering the relative distances of two competing genes from the LCR leads to changes in the levels of gene expression through alteration of the frequency of LCR-gene interactions (Dillon et al., 1997; this thesis).

We subsequently studied the role of erythroid Krüppel-like factor (EKLF) in gene competition between the human γ- and β-gene for LCR activation in compound knockout and human β-globin locus transgenic mice. These studies served to further illuminate the mechanism of developmental regulation. The zinc finger gene EKLF is a critical transcription factor essential in the activation of the murine adult β-globin genes through binding to the CACC element in both promoters (Nuez et al., 1995; Perkins et al., 1995). In the human β-globin locus a canonical EKLF binding site is found in the adult β-globin gene promoter (Feng et al., 1994) and additional binding sites are found in the hypersensitive sites of the LCR, particularly in the core region of HS site 3 (Philipsen et al., 1990; GT-rich sequences see Figure 4). Homozygous EKLF null fetuses are severely anaemic and die around day 15 due to a lack of adult β-globin gene expression whereas α-globin gene expression is unaffected. Expression analysis in EKLF+/- fetuses during stages of γ- to β-gene switching showed a decrease in the level of β-globin gene expression and reciprocal increase in the level of γ-gene expression which are fully accounted for by changes in the number of transcriptionally active genes (RNA FISH results in Chapter 3). Furthermore, increasing the levels of EKLF leads to an increase in the level of β-globin gene expression concomitant with reciprocal down regulation of γ-gene expression (Chapter 4). When competition is abrogated around day 16 of gestation due to the silencing of the γ-genes, expression of the β-gene returns to almost normal levels implying that the level of EKLF is most critical when the β-gene is in competition for the interaction with the LCR. In EKLF-/- fetuses the level of γ-gene expression is further increased while no transcriptionally active β-genes were detected. In addition, reciprocal changes in the sensitivity of the chromatin structure of the γ- and β-gene

promoters to DNaseI are observed. Similar results were obtained in expression studies in which EKLF null mice were bred with mice harbouring a human β-globin yeast artificial chromosome (Yac) transgene. In those studies only EKLF-/- mice were examined which showed a dramatic reduction in the expression of the β-globin gene while γ-globin transcripts were elevated about 5 fold (Perkins et al., 1996). This data is compatible with the competitive model of gene switching and implies that EKLF through binding to the CACC box mediates an adult stage specific interaction between the β-gene promoter and the LCR excluding the interaction with the γ-gene. Decreasing levels of EKLF affect the stability of LCR-β-gene interactions allowing the LCR to interact more frequently with the γ-genes. These observations are in full support of the looping model. The fact that down regulation of the most 3' adult β-globin gene results in reciprocal increase of the upstream γ-genes are more difficult to reconcile with the accessibility as well as the tracking model. However, these experiments do not completely exclude the possibility that either one of the alternative mechanisms can operate in the locus. This will eventually require more direct proof such as the visualisation of the dynamic interactions of the LCR with the individual globin genes in single erythroid cells during periods of gene switching. The EKLF data confirms once more that gene expression in the human β-locus is an all or none event suggesting that the number and length of the transcriptional active periods of the individual genes determine the level of gene expression within a single red cell. In the looping model the number and length of transcriptional active periods are translated into frequency and stability of LCR-gene interactions, respectively. These studies indicate, as was predicted before, that transcription factors are important regulators of the stability of LCR-gene interactions (Grosveld et al., 1993). In a parallel study we have shown that the relative distance of the genes from the LCR determines the frequency of interactions (Dillon et al., 1997).

*Primary RNA transcript detection*

FISH techniques have been established over the past decade for a variety of molecular genetic and cell biological applications (reviewed in Joos et al., 1994; Swiger and Tucker, 1996; Heng et al., 1997). In the work presented here, we have developed and applied an indirect RNA fluorescent *in situ* hybridisation procedure with the objective to detect primary transcripts at the site of synthesis in interphase nuclei of erythroid cells. With the use of this technique we obtained some valuable insights into the dynamics of gene transcription in single human β-globin loci. Gene specific primary transcripts were detected by applying four 50bp intron specific sequences containing biotin, digoxigenin or dinitrophenol (DNP) side chains in the middle and on the 5' and 3' sides. Neighbouring haptens were spaced with a minimum of 25bp to allow an optimal detection using a triple layered immunodetection system with fluorescently labelled antibodies or the biotin/(strept)avidine detection system. The major advantage of using indirectly labelled probes is the ability to amplify the signal of interest. Another major advantage of RNA FISH is the combinatorial use of different haptens and a variety of fluorochromes which allows for the visualisation of different nucleic acid sequences simultaneously as well as the codetection of proteins.

The results we have obtained are highly reproducible with little variation between slides of the same specimen and only a slightly higher variance between slides of different specimens, though on average this never exceeds 5%. The probe penetration and detection liability are almost complete as over 95% of double α-gene signals can be detected (Chapters 2 and 5). The use of RNA FISH in combination with DRB induced transcriptional inhibition and release also gave highly reproducible results even though the technical approach of this type of experiment is far more demanding.

The use of a multi-layered immunodetection system amplifies our primary hybridisation signal to reach the sensitivity needed to visualise their accumulated localisation at the site of

the gene. However, there are limits to the numbers of layers that can be added before non-specific annealing (background) starts to overwhelm the signal of interest. This background appears at a variable density and intensity as grainy speckles throughout the nucleus and is predominantly the result of aspecific sticking of probe sequences rather than the immunoreagents. Although the background hardly ever exceeds unacceptable levels in conventional RNA FISH experiments with the current number of amplifications, they do sometimes tend to become a problem in more difficult experiments. For example when combined with procedures such as cell sorting, cell culturing and enzymatic or otherwise more radical treatments. Most probably, these additional challenging steps may lead to the loss of target RNA sequences at its original site *in vivo* or to a general malaise of the target cells which may cause the loss of gene transcription. Thus additional layers of detection or a more sensitive detection system may guarantee a higher consistency in the results of more demanding procedures.

Substantial improvement in the signal to background ratio can be achieved through higher stringency of hybridisation and washing procedures. However this doesn't always lead to the expected results mostly because the intensity of specific signals also diminish as a result of the higher stringency. Increasing the number of reporter molecules at the site of hybridisation would be an alternative way to improve the signal to background ratio. This can be achieved through simply increasing the total length of probe sequences used in the hybridisation reaction. In the case of the human β-genes the entire intervening sequence can be used thereby increasing the total probe length almost five fold. The second intron of the adult human β-globin gene (850bp) has been labelled as a double stranded probe using nick translation to incorporate biotin or digoxigenin labelled dUTP, and after the application to fixed erythroid cells in RNA FISH experiments produced similar results (Gribnau and Milot, pers. comm.). The primary transcript signals had similar intensity to the signals produced using the four oligonucleotide sequences as probes, implying that in the case of the β-globin gene we have reached maximum intensities in their detection. This suggests that addition of more probe sequences may only be useful when the specific signals are at the border or just below the level of detection. However, when the amount of possible probe sequences that can be used in the hybridisation reaction is limited and insufficient, an alternative way is to use a more sensitive secondary detection system. Most promising in this respect is the use of a tyramide/peroxidase based ultrasensitive-detection technology, which mediates the deposition via covalent linkage of biotin or fluorochrome labelled tyramids at the site of hybridisation in a time dependent manner (Raap et al., 1995). Using this detection method we have been able to visualise the transcription of the targeted allele of the erythroid specific transcriptional activator EKLF in primitive and definitive erythroid cells using a nick transcription labelled 2.9kb LacZ probe (Nuez et al., 1995; Tewari, pers. comm).

The fluorescent *in situ* hybridisation detection of RNA seems to work best on highly transcribed genes such as the human and murine α- and β-like globin genes (this thesis), the murine *H19* and *Igf2* genes (results not shown), viral transcripts and the human β-actin gene (Zhang et al., 1994; Lawrence et al., 1989; Custodio et al., submitted). For a meaningful interpretation of the results it would be helpful to determine the sensitivity in absolute number of target RNAs that can be detected, which has not been achieved yet. The *in situ* hybridisation procedure is a multistep procedure and each step can add to the variability of the final measurement. In addition, our lack of knowledge about the stoichiometry of the hybridisation, the loss of target RNAs during the procedure and the efficiency of the secondary immunodetection system makes it even more difficult to make a proper assessment of the absolute target numbers. Still some idea of the sensitivity can be given. The average number of precursor molecules has been calculated to correspond to 110 β-globin molecules per adult erythroid cell nucleus (Kantor et al., 1980). Assuming that all globin genes have

similar transcription rates, an estimate of maximal 55 precursor RNA molecules are localised per allele. When only a single oligonucleotide instead of the usual four are used to detect the murine *βmaj* gene, signals are still visible as small pin-pricks in about 50% of the usual numbers of nuclear transcription signals, implying a minimal detection level of about 14 primary transcripts (results not shown). Alternatively, an erythroid cell accumulates about 20.000 to 25.000 globin mRNA molecules over a 60-72 hrs period during terminal differentiation (Clissold et al., 1977; Nienhuis and Benz, 1977). Taking the half life into account, which is between 24 and 48 hrs ((Ross and Sullivan, 1985; Stolle and Benz, 1988), twice at much has to be produced in order to come to this number of transcripts. This means a production of between 11.6 and 13.9 transcripts per min leading to a transcription initiation rate of every 8.6 to 10.4 seconds per allele (average of 9.5 sec), which is similar to the rate found in independent experiments in yeast studies (Iyer and Struhl, 1996). Transcriptional inhibition and release experiments at 37° (Gribnau, pers. comm) revealed that it takes 2.5 min for the reappearance of 50% of the normal numbers of fluorescent signals, corresponding to about 16 new initiation events. A rough estimation of the minimum detection sensitivity would be between 14 to 16 transcripts. Suggesting that when the hybridisation is complete this number of transcripts comes to a total length of 2800bp to 3200bp of haptenised probe sequences. This would predict that it should be possible to detect transcription of many eukaryotic genes. Indeed, with the use of RNA FISH studies 4.4kb and 6.5kb of probe sequences have been used to visualise primary transcription of the c-fos and fibronectin genes, respectively (Huang and Spector, 1991; Xing et al., 1993). Initiation taking place about every 9.5 seconds and an RNA polymerase II transcription rate of about 1250 nucleotides per min (Thummel et al. 1990; Shermoen and O'Farrell, 1991; Iyer and Struhl, 1996), suggests an average of 9 to 10 nascent transcripts on a transcriptionally active β-globin gene. This is less than our estimated detection limit and implies that we are normally detecting not only nascent transcripts but also intermediately processed precursor transcripts and probably some spliced intron lariats at the site of transcription. This seems plausible since precursor transcripts remain at the site of transcription till fully processed and only as mature mRNA are transported from the site of transcription towards the cytoplasm (Custodio et al., submitted). It can be expected that with further optimisation of the FISH procedure in combination with improvements in detection with digital imaging devices, the sensitivity and specificity of the *in situ* hybridisation on RNA targets will improve in the near future which should make it possible to detect the transcriptional activity of any single gene.

In recent years, FISH detection methods of primary RNA molecules in combination with immunocytochemical detection of nuclear proteins and digital imaging microscopy have been used to study in 3 dimensions the localisation and dynamics of the synthesis, processing and export of nuclear RNA (Xing et al., 1993; Carter et al., 1993; Zhang et al., 1994; Mistelli et al., 1997; Custodio et al., submitted). Early studies in NIH3T3 cells have visualised a direct link between the nuclear localisation of serum inducible *c-fos* transcription and nuclear speckles that are enriched in pre-mRNA splicing factors (Huang and Spector, 1991). In an attempt to better define the spatial relation between transcription and splicing, Zhang et al., 1994 showed for viral (Ad2) major late and the human endogenous β-actin gene that splicing occurs at the site of gene transcription. These studies imply that the nucleus is not divided into separate compartments regarding these nuclear processes.

Recently, we have explored the combinatorial use of RNA FISH to visualise globin gene transcription with immunodetection of nuclear localised proteins such as the proliferating cell nuclear antigen (PCNA). PCNA is highly conserved in eukaryotes and is essential for DNA replication and both DNA excision and mismatch repair (Kelman et al., 1997; Jonsson and Hubscher, 1997). The nuclear localisation of PCNA proteins varies during the different stages of the cell cycle. Visualisation of its nuclear distribution pattern using immunohistochemical

detection methods allows for the determination of the stage of the cell cycle of the individual erythroid cells. This will enable us to investigate the link between gene transcription and the cell cycle and may lead to better insights into the cell timing effect on gene transcription as observed and discussed by Milot et al., (1996).

Possible applications for the use of FISH detection of RNA may also be found for diagnostics in the clinic. Chromosome translocations in which a chimaeric gene product is being transcribed may be visualised as a combination of the two colours used to detect the RNA products of the individual genes. However, also more general transcription or processing defects may be visualised with this technique. Recently it has been shown that splicing mutations in the adult β-globin gene result in accumulation of transcripts at the site of transcription, providing a link between correct processing of the RNA and its transport from the site of transcription to the cytoplasm (Custodio et al., submitted).

*The role of EKLF in the erythroid lineage*

Erythroid Krüppel-like factor (EKLF) is a transcriptional activator first identified as an erythroid enriched cDNA by subtractive hybridisation (Miller and Bieker, 1993). EKLF contains three archetypal zinc fingers in its DNA binding domain which are structurally related to the Krüppel-family of transcription factors originally identified in *Drosophila*. With its zinc fingers it can recognise and bind the DNA consensus sequences CCA CAC CCT which has been shown to play a critical role in human β-globin gene expression but also in other related sequences like the β-LCR (Feng et al., 1994). Genetic ablation of both EKLF alleles in mice results in a lethal anaemia due to almost complete lack of adult β-globin gene expression and the mice die around day15 of gestation (Nuez et al., 1995; Perkins et al., 1995). Expression analysis of the role of EKLF in human globin gene expression transgenic mice showed an exclusive role in the regulation of the adult human β-globin gene (Chapter 3; Perkins et al., 1996). Primitive erythropoiesis was normal and expression studies in both primitive and definitive erythroid cells showed that non of the other β-like globin genes nor the α-like globin genes as well as many other genes with CACC elements in their promoters were affected (Nuez et al., 1995; Perkins et al., 1995; Wijgerde et al., 1996; Chapter 3; Perkins et al., 1996). We recently checked for the expression of two additional potential target genes called δ aminolevulinic acid synthase (ALA-S(E)) and ferrochelatase (FC), both enzymes involved in the heme synthesis pathway which like the globin genes become induced at the terminal stages of erythroid differentiation. However, as has been shown for the porphobilinogen deaminase (PBGD) gene, which is also involved in the synthesis of heme, also ALA-S(E) and FC were normally expressed in erythroid cells of EKLF null mice (results not shown).

Expression of EKLF proteins is not developmental specific and we have shown that EKLF functions as a transcriptional activator in embryonic erythroid cells (Chapter 4). Surprisingly, expression of the β-globin gene in primitive cells of EKLF-/- transgenic mice having a single copy of the μLCR-β-globin construct was still expressed and transcriptionally active in about 50% of the erythroid cells compared to wild type littermates (μD14 line in Ellis et al., 1996). In the definitive cells expression drops considerately and no active β-genes are detected (Chapters 4 and 5). This suggests that the adult β-gene is not fundamentally dependent on EKLF but only acquires this dependency in the definitive erythroid cells. A likely hypothesis may be that a different factor can or is allowed to substitute for EKLF on the β-globin CACC box during primitive stages of development when the conditions of the chromatin are less restrictive compared to definitive stages (Tewari et al., 1996). Possible candidates are Sp1 and BKLF both family members of the Krüppel family of transcription factors which are able to bind to CACC elements and which are expressed more abundantly than EKLF in the erythroid

lineage (Crossley et al, 1996). An alternative possibility might be that the juxtaposition of the β-gene to the LCR may under these milder chromatin conditions provide EKLF independency in some other ways. There is evidence that regions of the LCR are transcribed and having the β-globin promoter in close proximity might somehow alter the transcriptional potential or requirements of the juxtaposed β-gene (Tuan et al., 1992; Ashe et al., 1997). In addition, the murine β-genes which are normally expressed at about 10-15% in primitive erythroid cells, are inactive in EKLF-/- primitive erythroid cells (Perkins et al., 1995; Trimborn, pers. comm.). However the murine adult β-genes in these experiments are competing for LCR activity with the LCR more proximal embryonic εy- and βH1-globin genes which may cause a more dramatic loss of its expression compared to the non competing transgenic human β-gene.

We have shown that EKLF, rather than Sp1 or any other CACC box binding factor, interacts with one or more CACC motifs in HS site 3 of the β-LCR (Tewari et al., 1998; Chapter 4). However in the context of the full LCR EKLF seems to have a redundant function on the transcriptional activation and chromatin structure of the LCR (Chapters 3 and 4). This is supported by the observation that in EKLF-/- embryos the ε- and γ-genes are unaffected and that in EKLF-/- fetuses the transcriptionally active γ-genes even increase in numbers. If the LCR would have been affected in its function one would have expected a decrease in the expression rather than an increase.

At present only the adult β-globin genes have been identified as target genes of EKLF, and using a candidate gene approach we and other groups have excluded at least some putative target genes (see Nuez et al., 1995; Perkins et al., 1995; discussion above). This raises the question whether there are any additional target genes and, if so, how can we identify them? We have shown that α-globin gene expression is unaffected in EKLF-/- fetuses, implying that its promoter functions in an EKLF independent manner. Whether EKLF has additional target genes *in vivo* can be simply tested through introducing a transgenic construct in which expression of the adult murine β-globin gene is under the control of the α-promoter and the β-LCR. Both regulatory regions are expected to function, at least in part, in an EKLF null background. If the globin locus is EKLF's only target site, one would expect the EKLF-/- fetuses to be rescued, at least to some extent. However, if they die at similar stages one would expect that EKLF activates additional target genes in the definitive erythroid lineage. Finding those target genes can be done by differential screening for expression of genes between wild type and EKLF-/- definitive erythroid cells.

## References:


Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N.J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.*, **11**, 2494-2509.

Bickel, S. and Pirrotta, V. (1990). Self-association of the *Drosophila* zeste protein is responsible for transvection effects. *EMBO J.*, **9**, 2959-2967.

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

Carter, K.C., Bowman, D., Carrington, W., Fogarty, K., McNeil, J.A., Fay, F.S. and Lawrence, J.B. (1993). A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus. *Science*, **259**, 1330-1335.

Clissold, P.M., Arnstein, H.R.V., and Chesterton, C.J. (1977). Quantitation of globin mRNA levels during erythroid development in the rabbit and discovery of a new β-related species in immature erythroblasts. *Cell 11, 353-361.*

Crossley, M., Whitelaw, E., Perkins, A., Williams, G., Fujiwara, Y. and Orkin, S.H. (1996). Isolation and characterisation of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells

and selected other cells. *Mol. Cell. Biol.*, **16**, 1695-1705.
Custodio, N., Carmo-Fonseca, M., Hurst, J., Pereira, S., Grosveld, F. and Antoniou, M. (1998). Transport of mRNA from the site of transcription requires efficient splicing and 3'end processing. *Submitted.*
Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell.*, **1**, 131-139.
Ellis, J., Tan-Un, K.C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S. and Grosveld, F. (1996). A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human β-globin locus control region. *EMBO J.*, **15**, 562-568.
Enver, T., Raich, N., Ebens, A.J., Papayannopoulou, T., Costantini, F. and Stamatoyannopoulos, G. (1990). Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature*, **344**, 309-313.
Feng, W.C., Southwood, C.M. and Bieker, J.J. (1994). Analyses of β-thalassaemia mutant DNA interactions with erythroid Krüppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol. Chem.*, **269**, 1493-1500.
Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990). DNase I hypersensitive sites 1,2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucl. Acids Res.*, **18**, 3503-3508.
Fraser, P., Pruzina, S., Antoniou, M. and 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.
Furukawa, T., Zitnik, G., Leppig, K., Papayannopoulou, T. and Stamatonyannopoulos, G. (1994). Coexpression of γ and β globin mRNA in cells containing a single human β globin locus: Results from studies using single-cell reverse transcription polymerase chain reaction. *Blood*, **83**, 1412-1419.
Furukawa, T., Navas, P.A., Josephson, B.M., Peterson, K.R., Papayannopoulou, T. and Stamatonyannopoulos, G. (1995). Coexpression of ε, $^{G}\gamma$ and $^{A}\gamma$ globin mRNA in embryonic red blood cells from a single copy β-YAC transgenic Mouse. *Blood Cells Mol. Dis.*, **21**, 168-178.
Grosveld, F., Blom van Assendelft, G., Greaves, D.R. and Kollias, G. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell*, **51**, 975-985.
Grosveld, F., Dillon, N., and Higgs, D. (1993). The regulation of human globin gene expression. *Chapter 2 in "Baillière's Clinical Haematology"*, **6**, 31-55.
Groudine, M. and Weintraub, H. (1982). Propagation of globin DNase I -hypersensitive sites in absence of factors required for induction: A possible mechanism for determination. *Cell*, **30**, 131-139.
Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991). Importance of globin gene order for correct developmental expression. *Genes Dev.*, **5**, 1387-1394.
Heng, H.H.Q., Spyropoulos, B. and Moens, P.B. (1997). Fish technology in chromosome and genome research. *BioEssays*, **19**, 75-84.
Herendeen, D.R., Kassavetis, G.A. and Geiduschek, E.P. (1992). A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science*, **256**, 1298-1303.
Huang, S. and Spector, D.L. (1991). Nascent pre-mRNA transcripts are associated with nuclear regions enriched in splicing factors. *Genes Dev.*, **5**, 2288-2302.
Iyer, V. and Struhl, K. (1996). Absolute mRNA levels and transcriptional initiation rates in *saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*, **93**, 5208-5212.
Joos, S., Fink, T.M., Ratsch, A. and Lichter, P. (1994). Mapping and chromosome analysis: The potential of fluorescence in situ hybridisation. *J. BioTech.*, **35**, 135-153.
Kantor, J.A., Turner, P.H. and Nienhuis, A.W. (1980). Beta Thalassaemia: Mutants which affect processing of the beta-globin mRNA precursor. *Cell*, **21**, 149-157.
Lawrence, J.B., Singer, R.H. and Marselle, L.M. (1989). Highly localised tracks of specific transcripts within interphase nuclei visualised by in situ hybridisation. *Cell*, **57**, 493-502.
Martin, D.I.K., Fiering, S. and Groudine, M. (1996). Regulation of β-globin gene expression: straightening out the locus. *Curr. Opin. Genes Dev.*, **6**, 488-495.
Miller, I.J. and Bieker, J.J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol. Cell. Biol.*, **13**, 2776-2786.
Milot, E., Stouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. and Fraser, P. (1996). Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell*, **87**, 105-114.
Misteli, T., Caceres, F. and Spector, D.L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature*, **387**, 523-527.
Muller, H.P., Sogo, J.M. and Schaffner, W. (1989). An enhancer stimulates transcription in *trans* when attached to the promoter via a protein bridge. *Cell*, **58**, 767-777.
Muller, H.P. and Schaffner, W. (1990). Transcriptional enhancers can act in trans. *Tig*, **6**, 300-304.
Nienhuis, A.W., and Benz, E.J.Jr. (1975). Regulation of hemoglobin synthesis during the development of the red

cell (First of Three parts). N Engl. J. Med. 297, 1318-1328.
Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, **375**, 316-318.
Perkins, A.C., Sharpe, A.H. and Orkin, S.H. (1995). Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, **375**, 318-322.
Perkins, A.C., Gaensler, K.M.L. and Orkin, S.H. (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. USA*, **93**, 12267-12271.
Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990). The β-globin dominant control region: hypersensitive site 2. *EMBO J.*, **9**, 2159-2167.
Ptashne, M. (1988). *Nature*, **335**, 683-689.
Raap, A.K., van der Corput, M.C.P., Vervenne, R.A.W., van Gijlwijk, R.P.M., Tanke, H.J. and Wiegant, J. (1995). Ultra-sensitive FISH using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. *Hum. Mol. Gen.*, **4**, 529-534.
Ross, J., and Sullivan, T.D. (1985). Half-lives of beta and gamma globin messenger RNAs and of protein synthetic capacity in cultured human reticulocytes. *Blood 66, 1149-1154.*
Shermoen, A.W. and O'Farrel, P.H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell*, **67**, 303-310.
Stolle, C.A., and Benz, E.J.Jr. (1988). Cellular factor affecting the stability of beta-globin mRNA. *Gene 62, 65-74.*
Strouboulis, J., Dillon, N. and Grosveld, F. (1992). Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev.*, **6**, 1857-1864.
Swiger, R.R. and Tucker, J.D. (1996). Fluorescence in situ hybridisation: a brief review. *Env. Mol. Mutag.*, **27**, 245-254.
Tewari, R., Gillemans, N., Harper, A., Wijgerde, M., Zafarana, G., Drabek, D., Grosveld, F. and Philipsen, S. (1996). The human β-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene. *Dev.*, **122**, 3991-3999.
Tewari, R., Gillemans, N., Wijgerde, M., von Lindern, M., Grosveld, F. and Philipsen, P. (1998). Erythroid Krüppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the β-globin locus control region. *EMBO J., in press.*
Tuan, D., Kong, S. and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc. Natl. Acad. Sci. USA*, **89**, 11219-11223.
Wijgerde, M., Grosveld, F. and Fraser, P. (1995). Transcription complex stability and chromatin dynamics *in vivo*. *Nature*, **377**, 209-213.
Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F. and Fraser, P. (1996). The role of EKLF in human β-globin gene competition. *Genes & Dev.*, **10**, 2894-2902.
Xing, Y., Johnson, C.V., Dobner, P.R. and Lawrence, J.B. (1993). Higher level organisation of individual gene transcription and RNA splicing. *Science*, **259**, 1326-1330.
Zhang, G., Taneja, K.L., Singer, R.H. and Green, M.R. (1994). Localisation of pre-mRNA splicing in mammalian nuclei. *Nature*, **372**, 809-812.

## Summary

The genetic information that is required for the development of a single fertilised egg into an adult human body is stored as DNA in the nucleus of each cell. Genes are regions of DNA which encode for proteins that constitute the fundamental building blocks of a cell and each cell of the human body contains about 100.000 different genes. A large number of genes are expressed in all cells whereas expression of a smaller number of genes is restricted to specific cell types. Correct gene expression demands precise mechanisms of gene regulatory control, which is at the base of all biological processes.

The expression of many genes is predominantly regulated at the level of initiation of DNA transcription, and the human β-globin locus has been used as a versatile paradigm for the study of stage and tissue specific regulation of gene expression. The human β-globin locus contains embryonic (ε), fetal ($^{G}\gamma$, $^{A}\gamma$) and adult (δ, β) specific genes arranged in the same order as they are expressed during development. Two major switches in the expression of the β-globin genes occur during development, one from embryonic ε- to the fetal γ-globin genes and the second switch from fetal γ- to adult δ- and β-globin genes. For proper regulation, all the β-globin genes are dependent on the presence of an important regulatory sequence called the locus control region (LCR). The LCR is a 20kb DNA region which is located upstream of the genes and whose regulatory features are contained within five erythroid specific DNaseI hypersensitive regions: 5' LCR (HS5,4,3,2,1)-ε-$^{G}\gamma$-$^{A}\gamma$-δ-β 3' (see Figure 3, page 15). Genetic studies have indicated that during development all the genes in the locus are in competition for the activation function of the LCR. Our current working model proposes that the LCR activates the different genes through looping out of intervening DNA sequences establishing direct LCR-gene interactions. However, two alternative models have been proposed, and thus the question of how the LCR activates the genes over large distances still remains a topic of debate (see Figure 7, page 45). In the work presented in this thesis, we have explored β-globin gene switching in transgenic mice to distinguish between the different transcriptional activation mechanisms.

Gene competition suggests that only a single gene can be activated at any one time. In Chapters 2 and 5, we have addressed this question by using fluorescent *in situ* hybridisation with gene specific intron sequences as probes to visualise short lived events at the site of transcription in single erythroid cell nuclei. Signals appear as bright fluorescent foci at the site of the gene reflecting active gene transcription or very recent transcriptional activity. The results in mice that are homozygous for the transgenic human locus reveal both the detection of single and double gene signals; however, single gene signals predominate suggesting that the LCR activates only a single gene at a time. Furthermore the majority of switching cells (cells with different combinations of both gene signals) show that each globin locus can respond differently to the same transacting factor environment, suggesting that a dynamic activation mechanism operates in the locus. The double gene signals were controversial in that they could be explained as either a transition state in the transcriptional activity of two genes or as two genes that are simultaneously active. Using RNA FISH analysis in combination with DRB mediated transcriptional inhibition and release during periods of γ- to β-gene switching, we show that the double gene signals represent the codetection of transcripts coming from two alternately initiated genes. These studies indicate that gene expression in the β-locus is an all or none event, and the number of transcriptionally active periods of a gene and the duration of such periods determines the level of gene expression within a single cell. In the looping model this implies that the level of gene expression is determined by how often (frequency) and for how long (stability) the LCR interacts with a gene.

The model predicted that the balance of γ- and β-gene transcription would gradually change during the switching period as a result of gradual changes in the transcription factor environment. In Chapters 3 and 4 we have tested this hypothesis by varying the concentrations of an erythroid specific transcriptional activator, erythroid Krüppel-like factor (EKLF), in compound knockout and human β-globin locus transgenic mice. EKLF binds preferentially to CACC elements found in the murine and human adult β-globin gene promoters. EKLF homozygous knockout mice (EKLF-/-) appear normal during embryonic stages, and the embryonic ε- and fetal γ-globin genes are expressed at normal levels even though EKLF is functional during these stages and seems to bind to sequences in the LCR HS3. In contrast, EKLF-/- mice develop a fatal anaemia as a result of severely reduced levels of adult β-gene expression, and die *in utero* around day 15 of gestation. Expression analysis in heterozygous EKLF knockout fetuses (EKLF+/-) showed a reduced level of β-gene mRNA with a reciprocal increase in the level of γ-globin mRNA. Changes in gene expression are a direct consequence of decreased numbers of transcriptionally active β-globin genes and a reciprocally increased number of transcriptionally active γ-globin genes which in turn determine the expression level of each of the genes. Overexpression of EKLF results in an opposite effect. In addition, changes in the sensitivity of the chromatin structure are detected at the adult β-globin gene promoter, which in EKLF-/- fetal erythroid cells becomes insensitive to DNaseI, whereas the γ-gene promoters increase in sensitivity. This data supports a role for EKLF in establishing and/or stabilising LCR/ β-gene interactions and confirms the hypothesis that gene switching is effectuated by subtle changes in the transcription factor environment. Surprisingly, we show that the adult β-globin gene is not intrinsically dependent on the presence of EKLF since it is expressed in EKLF-/- primitive cells from a single copy μLCR-β-globin transgene but acquires this dependency only in the definitive erythroid lineage. Yet, how this dependency is attained is not known.

We conclude that gene competition in the β-locus is a result of single gene transcription periods which alternate in a dynamic fashion during stages of gene switching. This implies that the LCR activates only a single gene at a time, most likely through direct interactions with the promoter regions of the genes, suggesting that these interactions are transient and dynamic. The level of gene expression is determined by how often and how long a gene is transcribed rather than modulating the rate of transcription. A progressive switch form γ- to β-gene transcription is thought to be the result of gradual changes in the transacting factor environment, causing a gradual shift from initially more γ-gene transcriptional periods towards more β-gene transcriptional periods and finally almost exclusively β-gene transcription.

## Samenvatting

De genetische informatie die nodig is voor de ontwikkeling van een bevruchte eicel tot een volwassen individu ligt als DNA opgeborgen in de kern van iedere cel in ons lichaam. Het DNA is opgedeeld in genen, die coderen voor eiwitten die de functionele bouwstenen vormen. De genetische informatie in iedere cel van de mens bestaat uit zo'n 100.000 genen waarvan een groot deel gebruikt wordt in iedere cel en een kleiner deel alleen gebruikt wordt in specifieke celtypen. Dit vereist een nauwkeurig regulatie-mechanisme, dat dan ook de basis vormt van alle biologische ontwikkelingsprocessen en essentieel is voor de reactie van cellen op invloeden uit de omgeving.

Het tot uitdrukking brengen (expressie) van vele voor eiwit coderende genen wordt hoofdzakelijk gereguleerd op het niveau van de initiatie van het afschrijven van het gen (DNA transcriptie). De β-globine locus van de mens staat vaak model voor de studie naar de weefsel specifieke regulatie van transcriptie tijdens de ontwikkeling. De β-globine locus bevat 5 verschillende genen die gerangschikt liggen in dezelfde volgorde als waarin ze tijdens de ontwikkeling tot expressie komen: epsilon(ε)-globine gen in het embryo, G-gamma($^{G}\gamma$)- en A-gamma($^{A}\gamma$)-globine genen tijdens de foetale periode en delta(δ)- en beta(β)-globine genen na de geboorte. Twee belangrijke wisselingen (switchen) in de expressie van de β-globine genen vinden plaats tijdens de ontwikkeling, één van het embryonale ε-gen naar de foetale γ-genen en de tweede van de foetale γ-genen naar de volwassen δ- en β-genen. De "locus control region" (afgekort LCR) is een regulerend DNA element dat gelegen is voor de vijf β-globine genen en onontbeerlijk is voor de juiste expressie van deze genen. De activerende functies van de LCR zijn gelokaliseerd in vijf rode bloedcel specifieke DNaseI hypergevoelige (HS1-5) gebiedjes: 5' LCR (HS5,4,3,2,1)-ε-$^{G}\gamma$-$^{A}\gamma$-δ-β 3' (zie Figuur 3, blz. 15). Genetische studies hebben aangetoond dat tijdens de ontwikkeling alle β-genen in het cluster in competitie zijn met elkaar voor de activerende werking die uitgaat van de LCR. In ons huidige werkmodel stellen we dat de LCR de verschillende genen activeert door het vormen van directe interacties met de genen waarbij het tussenliggende DNA als een lus naar buiten steekt. Dit model staat echter lijnrecht tegenover twee alternatieve modellen die zijn voorgedragen (zie Figuur 7, blz. 45), waardoor de vraag hoe de LCR de verschillende genen activeert nog steeds een onderwerp van discussie is en blijft. Voor het in dit proefschrift gepresenteerde werk, hebben we onderzoek verricht naar het mechanisme dat betrokken is bij de rode bloedcel specifieke activatie en het switchen van gen expressie van de genen in de β-locus, zodat we uitspraken kunnen doen over het transcriptie activatie mechanisme door welke de genen in de β-locus worden gereguleerd.

Competitie tussen de genen suggereert dat slechts één enkel gen actief kan zijn op elk willekeurig moment. In experimenten beschreven in de Hoofdstukken 2 en 5, hebben we deze vraag beantwoord door gebruik te maken van een fluorescentie *in situ* hybridisatie techniek en gen-specifieke intron sequenties als probes om kort levende gebeurtenissen op de plaats van transcriptie in de kern van individuele cellen zichtbaar te maken. Signalen verschijnen als heldere fluorescerende spotjes op de plaats van een actief of vrij recentelijk actief gen. De resultaten in transgene muizen die een volledige β-globine locus van de mens bevatten laten zowel enkel- als dubbel-gen signalen zien; echter, de enkel-gen signalen hebben veruit de overhand wat er op wijst dat de LCR op elk willekeurig tijdstip slechts één enkel gen kan activeren. De meerderheid van cellen waarin het switchen van genen plaatsvindt (cellen met verschillende combinaties van gen-signalen) laten zien dat iedere globine-locus verschillend kan reageren op dezelfde factor-omgeving wat duidt op een dynamisch activatie mechanisme in de regulatie van β-globine expressie. De dubbel-gen signalen waren omstreden omdat zij zowel uitgelegd konden worden als een overgangssituatie in de activatie van het ene gen naar

het andere gen dan wel dat beide genen tegelijkertijd actief konden zijn. Met behulp van de fluorescentie *in situ* hybridisatie techniek in combinatie met DRB gemedieerde remming en het vervolgens vrijgeven van transcriptie tijdens perioden dat er gewisseld wordt van γ- naar β-gen expressie, hebben we aangetoond dat de dubbel-gen signalen het gevolg zijn van de overgang in de activatie van transcriptie van het ene gen naar het andere gen. Deze studies maken duidelijk dat tijdens de switch van γ- naar β-gen expressie de genen beurtelings worden afgeschreven en dat het niveau van gen-expressie in een cel bepaald wordt door het aantal actieve perioden van een gen en de duur van deze perioden. In het looping model betekent dit dat gen-expressie wordt bepaald door hoe vaak (frequentie) en voor hoe lang (stabiliteit) de LCR-gen interacties duren.

Het looping model voorspelde dat de switch van γ- naar β-gen expressie veroorzaakt zou worden door langzame veranderingen in het milieu van regulerende factoren in de kern van de rode bloedcel. In Hoofdstukken 3 en 4 hebben we de rol van een rode bloedcel-specifieke transcriptie activerend eiwit "erythroid Krüppel-like factor" (afgekort EKLF) in de expressie van de β-globine genen bestudeert. EKLF is een factor die bij voorkeur bindt aan het CACC element in de promotor van het β-globine gen maar niet aan de soortgelijke elementen in de ε- en γ-genen. Genetische verwijdering van het EKLF gen (EKLF-/-) leidt tot afwezigheid van het EKLF eiwit in de rode bloedcel; echter, EKLF-/- muizen ontwikkelen zich normaal gedurende de embryonale/foetale fase. Expressie studies hebben laten zien dat de embryonale ε- en foetale γ-genen niet veranderen in hun expressie ondanks het feit dat EKLF eiwitten al actief zijn in het embryo en binden aan CACC elementen in HS3 van de LCR. EKLF-/- foetussen hebben een extreem laag expressie niveau van het volwassen β-gen. Deze lage β-expressie geeft aanleiding tot een ernstige bloedarmoede met dodelijke afloop rond de vijftiende dag van de draagtijd. Halvering van de EKLF concentratie in EKLF+/- muizen geeft aanleiding tot verminderde β-expressie en een wederzijdse verhoging in γ-gen expressie. De veranderingen in gen-expressie zijn een direct gevolg van een verandering in het aantal actieve γ- en β-genen welke recht evenredig zijn met het niveau van expressie. Het verhogen van de EKLF concentratie in de kern heeft het tegenovergestelde effect. Verder zijn er veranderingen in de gevoeligheid van de chromatine structuur waargenomen in de promotor van het β-gen welke in de rode bloedcellen van EKLF-/- foetussen minder toegankelijk wordt voor het enzyme DNaseI terwijl de promotors van de γ-genen juist toegankelijker worden. Deze resultaten duiden op een rol van EKLF in het tot stand brengen en het stabiliseren van LCR-β gen interacties en bevestigen daarmee de hypothese dat het switchen van de expressie van genen in de β-globine locus wordt veroorzaakt door veranderingen in het transcriptiefactor milieu. Echter, het volwassen β-globine gen is niet fundamenteel afhankelijk van EKLF, daar het β-gen tot expressie komt in de embryonale rode bloedcellen van EKLF-/- muizen transgeen voor het μLCR-β-gen construct. Het β-gen verwerft de afhankelijkheid van EKLF pas in de volwassen rode bloedcel.

Wij concluderen dat gen competitie in de β-globine locus een resultaat is van wisselende periodes waarin slechts één gen actief is. Dit suggereert dat de LCR op elk willekeurig tijdstip slechts met één gen een interactie aangaat, en deze interacties zijn tijdelijk en dynamisch. Het niveau van gen expressie wordt bepaald door de frequentie en stabiliteit van LCR-gen interacties. De switch van γ- naar β-gen expressie is geleidelijk en wordt veroorzaakt door veranderingen in het transcriptiefactor milieu die resulteren in een overgang van overwegend γ-gen transcriptie tot een geleidelijke vermindering van γ- en vermeerdering van β-gen transcriptie, tot uiteindelijk overwegend β-gen transcriptie.

**Curriculum vitae**

| | |
|---|---|
| Naam: | Marcus Govert Johannes Maria Wijgerde (Mark) |
| Geboren: | 24 oktober 1966 te Breda. |
| 1979-1987 | HAVO en VWO op het Dr. Mollercollege te Waalwijk. |
| 1987-1988 | Militaire dienst. |
| 1988-1993 | Medische Biologie, faculteit der Geneeskunde, Rijks Universiteit te Utrecht. Doctoraal: Cum Laude. Stage-ervaring: Onderwerp: Karakteriseren van de expressie van twee homeobox bevattende genen, *Hox-2.3* en *Hox-2.4*, gedurende en vroeg na gastrulatie (7,5-9,5 dpc) in de muis, m.b.v. *In Situ* Hybridisatie. O.l.v. Dr. J. Deschamps, Hubrecht Laboratorium te Utrecht. Onderwerp: Het analyseren van het mechanisme van Genomic Imprinting in de muis. O.l.v. Prof. Dr. A. Surani, WELLCOME/CRC Institute of Cancer and Developmental Biology, Universiteit van Cambridge, Groot-Brittanië. |
| 1993-1998 | Een half jaar als assistent in opleiding (a.i.o.) in dienst van de Erasmus Universiteit Rotterdam (EUR), waarna in maart 1994 een aanstelling als onderzoeker in opleiding (o.i.o) in dienst van de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) volgde. Gedurende beide periodes werkzaam in de groep van Prof.dr F.G. Grosveld, afdeling Celbiologie en Genetica aan de medische faculteit van de Erasmus Universiteit Rotterdam, waar gewerkt werd aan het aldus verschenen proefschrift. Wetenschaps prijs: Promovendus prijs 1997 toegekend door de Nederlandse Genetische Vereniging (NGV) voor het zeer oorspronkelijk onderzoek naar het mechanisme van de expressie van het beta-globine locus. |

**List of Publications**

Deschamps, J., and **Wijgerde, M.** (**1993**). Two phases in the establishment of *HOX* expression domains. *Developmental Biology 156, 473-480.*

Koide, T., Ainscough, J., **Wijgerde, M.**, and Surani, M,A. (**1994**). Comparative analysis of *Igf2/H19* imprinted domain: Identification of a highly conserved intergenic DNase I hypersensitive region. *Genomics 24, 1-8.*

**Wijgerde, M.**, Grosveld, F., and Fraser, P. (**1995**). Transcription complex stability and chromatin dynamics *in vivo. Nature 377, 209-213.*

Milot, E., Strouboulis, J., Trimborn, T., **Wijgerde, M.**, de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F., and Fraser, P. (**1996**). Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell 87, 105-114, (4 October 1996).*

Tewari, R., Gillemans, N., Harper, A., **Wijgerde, M.**, Zafarana, G., Drabek, D., Grosveld, F., and Philipsen, S. (**1996**). The human *beta-globin* control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial *lacZ* gene. *Development 122, 3991-3999 (1996).*

**Wijgerde, M.**, Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F., and Fraser, P. (**1996**). The role of EKLF in human beta-globin gene competition. *Genes &Development 10: 2894-2902 (1996).*

Ashe, H. L., Monks, J., **Wijgerde, M.**, Fraser, P., and Proudfoot, N.J. (**1997**) Intergenic transcription and transinduction of the human β-globin locus. *Genes & Development 11: 2494-2509 (1997).*

Tewari, R., Gillemans, N., **Wijgerde, M.**, Grosveld, F., and Philipsen, S. (**1998**). Erythroid Krüppel-like factor (EKLF) is active in primitive and definitive erythropoiesis and required for the function of 5'HS3 of the β-globin locus control region. *EMBO Journal 17, 2334-2341.*

Gribnau, J., de Boer, E., Trimborn, T., **Wijgerde, M.**, Milot, E., Grosveld, F., and Fraser, P. Chromatin interaction mechanism of transcriptional control. *Submitted*

Gillemans, N., Tewari, R., Lindeboom, F., de Wit, T., **Wijgerde, M.**, Grosveld, F., and Philipsen, S. EKLF but not Sp1 is an activator of the β-globin Locus Control Region *in vivo. Submitted*

## Nawoord

Dit proefschrift zou onvolledig zijn zonder een klein woord van dank. Daarom met deze laatste woorden bedank ik alle collega's, voornamelijk de globine-collega's, die mij op welke manier dan ook hebben geholpen en/of hebben bijgedragen aan de succesvolle, leerzame en gezellige Rotterdamse O.I.O.-jaren. In het bijzonder bedank ik Peter Fraser en Sjaak Philipsen voor de leerzame begeleiding. Speciale dank gaat uit naar Frank Grosveld omdat wanneer zijn deur openstond ik altijd terecht kon voor steun, adviezen en discussies.