β-Globin gene regulation and chromatin structure

β-Globine gen-regulatie en chromatinestructuur

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 voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 30 juni 1999 om 11.45 uur

door

Joost Henk Gribnau

geboren te Hengelo

Promotie commissie

Promotor:	Prof. dr F.G. Grosveld
Overige leden:	Prof. dr B. Löwenberg Prof. dr E.A. Dzierzak Prof. dr C.P. Verrijzer
Co-promotor:	Dr P. Fraser

Omslag: Polymerase op DNA, Joop Polder 1999.

Dit proefschrift is tot stand gekomen binnen de vakgroep Celbiologie en Gentica van de Faculteit der Geneeskunde en Gezondheidswetenschappen van de Erasmus Universiteit te Rotterdam. De vakgroep maakt deel uit van het Medisch Genetisch Centrum Zuidwest Nederland.

Aan mijn ouders,
voor Susanne.

Contents

List of abbreviations

Bp	Base pair
CTE	Cell timing effect
DNA	Deoxy-nucleic acid
DRB	5,6-dichloro-1-β-D-ribofuranosylbenzimidazole
EKLF	Erythroid Kruppel-like factor
GTF	General transcription factor
HAT	Histone acetyl transferase
HbA	Adult hemoglobin
HbF	Fetal hemoglobin
HEL	Human erythroid cell line
HS	Hypersensitive site
IVS	Intervening sequence (intron)
Kb	Kilo base pairs
KD	Kilo Dalton
K562	Human erythroid cell line
LCR	Locus control region
MAR	Matrix attachment region
MEL	Murine erythroleukemia cell line
Mb	Mega base
mRNA	Messenger RNA
Nm	Nano meter
Pc	Polycomb
PE	Position effect
PEV	Position effect variegation
PRE	Polycomb responsive element
RNA	Ribonucleic acid
SAR	Scaffold attachment region
TBP	TATA binding protein
TF	Transcription factor
TRE	Trithorax responsive element
Trx	Trithorax
UTR	Untranslated region
YAC	Yeast artificial chromosome

Scope of this thesis

Oxygen transport in the blood is mediated by highly specialized red cells. The majority of the proteins of a red cell comprises the oxygen carrier protein hemoglobin, which is a hetero-tetrameric protein that consists of two α and two β globin chains. In humans different α- and β-like globin chains are expressed during development resulting in several different hemoglobin tetramers. The expression of different globin chains serves to facilitate the oxygen uptake by the embryo, since embryonic hemoglobin has a higher affinity for oxygen. The genes that code for the α- and β-like globin chains reside in loci located on separate chromosomes. Many blood disorders, like α- and β-thalassemias and sickle cell anemia are the consequence of deletions or mutations of sequences in these loci and they initiated extensive research into the molecular basis of these diseases. Especially the β globin locus has served as a model system to study the regulation of multi gene loci. The five functional globin genes, 5'-ε-^{G}γ-^{A}γ-δ-β-3', are differentially expressed during development. Proper expression of these genes requires the presence of a region located 5' upstream of the ε globin gene. This locus control region (LCR) contains 5 small regions that bind several trans-acting factors in an erythroid environment. The aim of this thesis is to study the role of the LCR in the regulation of the β globin genes *in vivo*.

Chapter 1 reviews the current knowledge in the regulation of eukaryotic transcription and chromatin. **Chapter 2** gives a broad introduction covering two decades of studies concerning the regulation of the human and murine β globin genes and serves as an outline for the results that will be discussed in the chapters 3 to 6. **Chapter 3** describes the role of EKLF in the γ to β globin switching process. **Chapter 4** and **5** describe experiments that show that the human and the murine LCR can only activate one globin gene at a given moment. **Chapter 6** describes the characterization of intergenic transcription in the human β globin locus. Finally in **chapter 7** the implications of the results presented in this thesis will be discussed.

Chapter 1

The regulation of eukaryotic transcription and chromatin structure.

An introduction

Eukaryotic Transcription

Transcription is a complex process used by organisms to copy DNA sequences into RNA that is usually translated in proteins. The transcription process in prokaryotic organisms is tightly coupled with translation and both processes can take place simultaneously. Prokaryotes utilise only one RNA polymerase for the transcription of all active genes. In eukaryotes transcription and translation are uncoupled, transcription is restricted to the nucleus whereas translation takes place in the cytoplasm. In contrast to prokaryotes, transcription in eukaryotes is mediated by three different RNA polymerases. RNA polymerase I is specifically used for the transcription of the ribosomal genes which are organised in tandem arrays on several chromosomes. Polymerase II is essentially used for transcription of all the genes encoding mRNA, small nuclear RNA and non coding RNA's as will be discussed in chapter 6. RNA polymerase III is utilised for the transcription of tRNA genes as well as the ribosomal gene of the 5S rRNA. In the context of this thesis I will focus on RNA polymerase II specific transcription and the factors, elements and processes involved in initiation, elongation and termination of transcription.

Promoters

Transcription initiation is a process in which many different trans-acting factors act on different DNA elements *in cis*. The TATA box and the initiator, separated by 25-30 base pairs, (Smale and Baltimore 1989) are the two core promoter elements, found upstream of most eukaryotic genes. Although not all genes have a TATA box element, for most genes, including the globin genes, this element is essential for gene activation. The TATA binding protein (TBP) specifically binds the TATA element and forms the basis of the TFIID complex that contains several TBP associated factors (TAF's). Like TFIID several other general transcription factors (GTF's) are part of the basic transcription machinery (TFIIA, TFIIB, TFIIE, TFIIF, TFIIH, reviewed in McKnight 1996). In cell free systems the TATA box plus the initiator element can drive transcription upon addition of extracts containing these general components. However, *in vivo*, in a chromatin environment, these elements are not sufficient to initiate transcription (Imbalzano 1994), indicating that additional elements are required for the proper activation of genes *in vivo*. Parts of these elements are promoter proximal elements, situated 50 to 200 base pairs upstream of the start site. Promoter proximal elements consist of typical recognition sites like CCAAT and CACC boxes that bind ubiquitous as well as specific trans-acting factors. Recent data show that many trans-acting factors, that bind these elements, can recruit chromatin-remodelling complexes and open up the chromatin of the promoter. This process of chromatin remodelling has been shown to be independent of the binding of the basic transcription machinery (Armstrong 1998).

Enhancers

Enhancers are DNA elements with a size between 50 base pairs to 1.5 kilo bases, situated up to several kilo bases upstream, downstream or within a gene, and are defined as elements providing increased transcription of a gene independent of the position or orientation relative to the transcription start site (Banerji 1981, Moreau 1981). Like promoter proximal elements, enhancer sequences provide additional specificity to the gene they activate by binding developmental, cell type or cell cycle specific factors. Three different models have been postulated to explain the mechanism by which an enhancer activates a gene. In the tracking model factors use

the enhancer as an entry site and track along the DNA in search of a gene (Herendeen 1992). In the accessibility model the enhancer provides a specific environment, like an open chromatin structure, necessary for optimal activation of a linked gene (Martin 1996). The looping model explains activation via direct interactions between gene proximal elements and the enhancer by looping out intervening DNA (Ptashne 1988, Mueller-Storm 1989, this thesis). The same models have also been put forward to explain the activation mechanism of the human β globin locus control region (LCR), which will be discussed in the next chapter. A lot of evidence in favour of a looping model has been published in recent years using the human β globin locus as a model system. Additional observations like transvection between chromosomes found in *Drosophila* (Henikoff 1997) and experiments showing looping *in vitro,* visualised with electron microscopy (Su 1991) seem to support these findings.

Transcription initiation

The largest subunit of the RNA polymerase II core enzyme contains a large carboxy-terminal domain (CTD) with 17-52 tandem repeats of a consensus heptapeptide. The CTD is essential for *in vivo* function of polymerase II. It is rich in potential phosphorylation sites that are hyperphosphorylated during transcription elongation (polIIO isoform) and hypophosphorylated prior to promoter clearance (polIIA isoform). Several factors have been purified that associate with the CTD, like SRB and mediator proteins and the general transcription factors TFIIB, TFIIF and TFIIH (Koleske and Young 1994). TFIIH, TFIIE and SRB proteins are CTD kinases and SRB and TFIIH show cyclin dependent kinase (CDK) activity, and link transcription initiation to cell cycle progression (for review see McKnight 1996). The discovery of the RNA polymerase II holo-complex in yeast suggests that RNA polymerase II binds the DNA as a holo-complex rather than being assembled on the

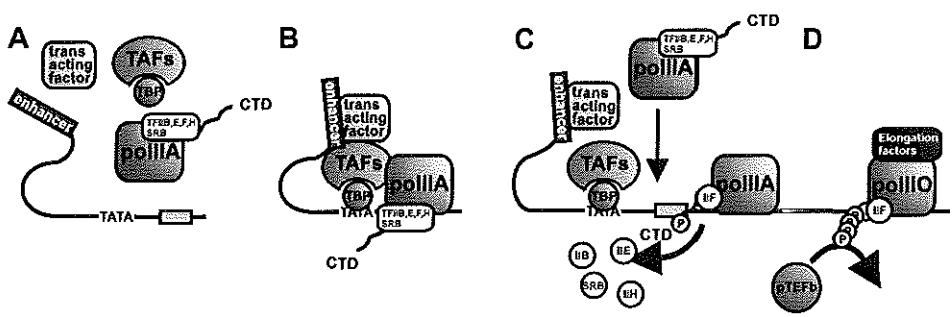


Figure 1. *A model for transcription initiation in eukaryotic organisms.Different transacting factors, TFIID and the polymerase II holocomplex bind enhancer, promoter proximal elements, TATA box and initiator respectively (A and B). Partial phosphorylation of the CTD and release of different CTD associated factors triggers promoter clearance followed by rapid reinitiation of new polymerase II holocomplex or dissociation of the different factors of the template (C). Polymerase II enters the late elongation phase after pTEF-b mediated hyperphosphorylation of the CTD (D).*

template. RNA polymerase II holo-complexes have also been isolated in mammals, indicating that the transcription initiation mechanism is highly conserved in eukaryotes (Ossipow1995). Promoter clearance is accompanied with the dissociation of the CTD associated factors TFIIB, TFIIH and the SRB proteins (Zawel 1995). However it is not clear if the dissociation of these factors is the key step that drives promoter clearance or if phosphorylation of the CTD precedes the dissociation of the different factors. Since hyperphophorylation of the CTD resulting in the polIIO isoform of polymerase II takes place after clearance of the promoter (Marshall 1996), partial phosphorylation could be the trigger to fire the polymerase.

Recent data of Struhl and colleagues show that autonomous transcription factor activation domains exert their function when connected to enhancer-bound proteins but fail to do so when linked to basic components of the polymerase II transcriptional machinery (Keaveney 1998). This shows the importance of enhancer recruitment of the polymerase II holo-complex and argues strongly in favour of a model with three basic complexes, the enhancer bound trans-acting factors, TFIID and the polymerase II holo-enzyme (Struhl 1996). The different complexes act synergistically via protein-protein interactions and recruitment of one complex increases the change of recruitment of another complex. The simultaneous binding of all three complexes results in promoter clearance. Subsequently the different complexes can dissociate from the promoter or rapid reinitiation can take place by attracting a new polymerase II holo-enzyme to the promoter (figure 1). It has been shown in yeast and *Drosophila* that highly active promoters are capable of initiating transcripts every 6 seconds, resulting in a polymerase II density of 1 polymerase per 100 base pairs (Struhl 1996).

Transcription elongation

After promoter clearance the RNA polymerase enters the elongation phase which can be separated into an early elongation phase, with a hypophosphorylated C terminal domain (CTD) of RNA polymerase II (polIIA), and a late elongation phase with a hyperphosphorylated CTD (polIIO). Hypophosphorylated polymerase II is blocked about 500 base pairs downstream of the transcription start site if phosphorylation fails to take place (Chodosh 1989). This was first shown by the use of DRB that specifically blocks the CTD kinase pTEF-b, a factor found in *Drosophila* and humans, and recently shown to be the cyclin dependent kinase CDK9 (Peng 1998). A sustained block of phosphorylation, which can be induced by DRB, results in the dissociation of stalled polymerases from the template giving rise to short prematurely terminated transcripts. Upon phosphorylation of the CTD the polymerase enters the late elongation phase and transcribes the genic sequences with a speed of around 1600 base pairs per second. Several factors found in yeast and in mammals (DmSII, TFII-F, ELL and Elongin, for review see Shilatifard 1997) are required for proper elongation. Although not shown to be required for elongation several other factors like SWI/SNF and pCAF are associated with polymerase II during elongation (Wilson 1996, Cho 1998). Interestingly the elongation mode seems to be dependent on the promoter that is used for initiation. The c-myc gene has two different promoters, one of which can give rise to blocked elongation whereas the other can not (Bentley 1986). Also cell cycle progression controls proper elongation since transcription elongation is aborted during mitosis (Shermoen and O'Farrell 1991). These and other data that show that the CTD is required for splicing (Corden 1997), mRNA capping (Cho 1997) and poly-adenylation (Dantonel 1997,

McCracken 1997), indicate that the process of transcription is tightly regulated and highly interactive.

Transcription termination

The rRNA specific RNA polymerase I terminates transcription in a specific region downstream of the rRNA gene (for review see Reeder 1997). For RNA polymerase II transcription such defined termination sequences do not seem to exist. Studies on the β globin genes show that the polymerase density decreases gradually behind the poly-adenylation signal, but remains detectable up to 500 base pairs downstream of the poly-adenylation sequence (Ashe 1997). Disruption of transcription termination after the deletion of different factors involved in poly-adenylation shows a clear correlation between poly-adenylation and transcription termination (Birse 1998) suggesting that the poly-adenylation signal is also an indirect transcription termination signal.

Chromatin

Cytological staining of eukaryotic cells reveals two different types of chromatin. Dark and light staining areas called heterochromatin and euchromatin respectively. The staining differences are attributed to differences in packaging of DNA. Most of the active genes reside in euchromatin that is generally situated in the interior of the nucleus and replicates early in S phase of the cell cycle. Hetero-chromatin is mostly situated at the nuclear periphery and replicates late in S phase. To date the initial cytological black and white picture has changed in a more extensive view with different chromatin states being essential tools for regulating transcription activation.

Chromatin structure and organisation

Chromatin is basically organised in three different levels, the 10 nanometer fibre, the 30 nanometer solenoid fibre and chromatin loops. The 10 nanometer fibre is an array of nucleosomes with 145-147 base pairs of DNA wrapped around each nucleosome, resulting in a 'beads on a string' like structure. A nucleosome is a multi-protein complex that consists of 8 histones, containing two copies of each histone protein, H2A, H2B, H3 and H4. Binding of linker histones H1 and H5 results in a more compact 30 nanometer diameter fibre or solenoid (Allan 1986). The amino terminal tails of histones can be modified by phosphorylation and acetylation leading to changes in the binding affinity of the nucleosome for DNA resulting in different local chromatin states. In *vitro* packaging of naked DNA in nucleosomes can easily suppress transcription initiation (Imbalzano 1994, Lorch 1987), whereas elongating polymerase II is only partially suppressed (Izban 1992). Both effects can be overcome *in vitro* (Tsukiyama 1995) and *in vivo* (Almer 1986) by specific proteins, like elongation factors, or chromatin remodelling complexes capable of opening up chromatin templates. It has been found that nucleosomes are disrupted at the promoter *in vivo* and interestingly several general transcription factors have domains highly homologous with histone DNA binding domains. This suggests that the transcription initiation complex may at least in part, replace nucleosomes in order to locally open up the chromatin. Contradictory reports have been published concerning the mechanism of polymerase elongation in chromatin. Several papers show the complete displacement of nucleosomes upon passage of a polymerase, whereas other groups did not find any changes in structure of the chromatin template. The truth will

probably be found somewhere in between with a partial displacement of the nucleosome and a concomitant uncoiling and recoiling of the DNA during passage of the polymerase.

The highest level of organisation is the arrangement of chromatin in big loops up to several mega bases in size, as was shown by electron microscopy and distance measurements using fluorescent *in situ* hybridisation (FISH, Yokata 1995). Different loops are anchored to the nuclear scaffold via scaffold attachment regions (SAR) or matrix attachment regions (MAR) depending on the method of isolation. Both MAR and SAR regions appear to overlap, but so far no consensus sequence has been found except for irregular spaced tracks of A's and T's. These tracks are recognised by several proteins like histone H1 and topoisomerase II, the latter being a major protein of the nuclear scaffold. Recently SAR's have been implicated in chromatin remodelling by linking a SAR element to an immunoglobulin μ gene enhancer. Addition of a SAR element proximal of the enhancer increased the chromatin accessibility and demethylation of a T7 promoter located 1 kb downstream of the immunoglobulin μ gene enhancer (Jenuwein 1997). These data suggest that SAR's are not only involved in higher order chromatin organisation but also have a function in chromatin remodelling.

The nucleus appears to be organised in specific domains like heterochromatin, euchromatin and sub chromosomal domains. Late replicating heterochromatin including centromeres and telomeres is preferentially found in the nuclear periphery, whereas euchromatin is mainly found in the nuclear centre and is replicated early in S phase. This type of organisation and the presence of SAR/MAR elements suggest a rather static organisation of chromatin. However work of Csink and Henikoff (1996) shows that insertion of heterochromatin close or inside euchromatin leads to heterochromatinization of surrounding euchromatin with a concomitant change in the localisation of the element within the nucleus. Other studies using the centromere binding protein Cenp-B attached to GFP confirm this picture and clearly show centromeric movements (reviewed by Lamond 1998), implying a more dynamic regulation of chromatin formation.

Histone (de)acetylation

One of the major mechanisms used for local remodelling of chromatin is the modification of different lysine residues of histones by adding or removing acetyl groups of the N-terminal tails of histones. Acetylation neutralises the positive charged lysine residues thereby decreasing the affinity of the nucleosome for DNA. General hyperacetylation has been correlated with transcriptional active areas (Hebbes 1994) and is associated with transcription (Lee 1993, Vetese-Dadey 1996), whereas hypoacetylated areas correspond with transcriptionally silent areas, including the in-active X chromosome in female mammals (Turner 1992, Jeppesen and Turner 1993). Specific acetylation patterns have been described for euchromatin and heterochromatin (Turner 92). Heterochromatin in yeast and *Drosophila* is only acetylated at lysine residue 12 (K12) of histone H4 although this has not been found in human chromatin. Different combinations of acetylated H3 and H4 lysine residues are used in euchromatin in *Drosophila*, yeast and humans (Turner 1992, Clarke 1993, O'Neill 1995).

Acetylation is mediated by histone acetyl transferases (HAT's), subdivided in A and B type HAT's. Type A is a nuclear HAT involved in the activation of transcription whereas type B is a cytoplasmic HAT involved in nucleosome assembly

associated acetylation. The first nuclear HAT to be identified was *Tetrahymena* GCN5 (Brownell 1996), named after the yeast homologue that was earlier reported to be a transcriptional co-activator (Georgakopoulos and Thireos 1992). Yeast mutants show that GCN5 is required for gene activation. GCN5 preferentially acetylates K14 of H3 and K8 and K16 of H4 (Kuo 1996). GCN5 seems to act locally by targeting to the promoter and is found in at least two protein complexes, GCN5p/ADA and SAGA (Grant 1997). These complexes appear to act as a bridge between transacting factors and the general transcription machinery. GCN5 overexpression studies showed a general increased acetylation, although it has not been proven that this is a direct effect. Even though several HAT's have been characterised it is not yet clear how complete domains are acetylated as has been found in the chicken β globin locus (Hebbes 1994). Several other known proteins are found to have histone acetyl transferase activity. Most of these proteins are directly involved in the transcription activation process, like human TAF250, a basic component of the general transcription factor TFIID (Mizzen 1996). CPB/p300 another well studied coactivator possesses HAT activity and is associated with the hypoacetylated isoform of polymerase II, which suggests that it plays a direct role in the transcription initiation process (Ogryzko 1996). In contrast with CBP/p300 another HAT pCAF appears to be associated with the hyperacetylated elongating polymerase II (Cho 1998), indicating that pCAF is possibly involved in acetylating histones of the transcribed template of active genes. More recently it has become clear that histones are not the only targets of HAT's since other proteins like general transcription factors can also be targets of acetylation. CBP/p300 for instance is capable of acetylating TFIIEβ, TFIIF and p53 *in vitro* (Imhof 1997). Two recent reports show that the activity of the erythroid specific transcription factors GATA1 and EKLF is also dependent on acetylation (Boyes 1998, Zhang 1998).

Histone acetylation is a reversible process and different deacetylases have been identified. The best-described deacetylases are yeast Rpd3 and its human homologue HDAC1. Rpd3 appears to interact with Sin3 and deletions of either protein give rise to identical phenotypes, whereas double mutants don't show additive effects. Mutant Rpd3 strains show increased acetylation of H4 lysine residues K5 and K12 and are defective in transcriptional repression *in vivo* (Kadosh and Struhl 1998a). Another study showed that like GCN5 the action of Rpd3 is highly localised to the promoter area, resulting in deacetylation of only one or two nucleosomes (Kodosh and Struhl 1998b). A second human deacetylase, HDAC2, associates with different DNA binding repressors like Mad (Hassig 1997), Ume6 (Kadosh 1997) and YY1 (Yang 1996). Interestingly a *Drosophila* homologue of YY1 called pleiohomeotic shows sequence specific binding to Polycomb responsive elements (Brown 1998).

Chromatin assembly

Upon entry of S-phase new nucleosomes have to be assembled on the replicating template. This assembly is tightly coupled with the proceeding replication fork that progresses at a speed of ~1000 base pairs per second (Lucchini 1995). After replication the parental nucleosomes randomly segregate over the parental and the newly synthesised strand (Sogo 1986) and open spaces are filled in with new nucleosomes. To date it is not clear how specific acetylation patterns are passed on to the replicated chromatin. Nucleosome assembly starts with binding of the H3/H4 tetramer to the DNA, a process facilitated by chromatin assembly factor 1 (CAF1),

followed by the assembly of the H2A/H2B dimers. Newly synthesised histone H4 is preferentially acetylated at lysine residues K5 and K12 in human and *Drosophila*, whereas the H3 acetylation pattern seems to be less conserved. Together with CAF1 these specifically acetylated H3 and H4 form the chromatin assembly complex (CAC, Verreault 1996). Shortly after incorporation in chromatin the newly synthesised nucleosome can be deacetylated (Jackson 1976).

Chromatin remodelling

Several factors directly involved in the transcription activation process, like the TATA binding protein (TBP) can not bind DNA if this specific sequence is occluded by a nucleosome. Chromatin remodelling complexes have been implicated in enhancing the process of activator binding, and repositioning of nucleosomes and nucleosome spacing after chromatin assembly. Both processes appear to be dependent on ATP hydrolysis. Several chromatin remodelling complexes (CRC's) have been purified recently from different organisms having one or both activities and the list is growing rapidly.

The first CRC identified was the yeast SWI/SNF complex, a multi-protein complex required for transcription of a specific subset of genes such as HO (Stern 1984) and SUC (Laurent 1990). The SWI/SNF complex contains 11 proteins (for review see Kadonaga 1998) of which ISWI, the *Drosophila* homologue of the yeast SWI/SNF subunit SWI2/SNF2, possesses ATP dependent remodelling activity (Corona 1999). Interestingly mutations in both the structured domain and N-terminus of histones appear to suppress the phenotype of SWI/SNF mutations, implying that the SWI/SNF complex is involved in overcoming repressive effects of chromatin (Recht 1999). SWI/SNF facilitates targeting of transcription factors to specific promoters and disrupts the local nucleosome structure. RSC was the second SWI/SNF like complex isolated from yeast (Cairns 1996). This complex is more abundant than SWI/SNF and seems to act on a broader range of genes. SWI/SNF related complexes have also been isolated from *Drosophila* (Varga-Weisz 1997, Tsukiyama 1995, Ito 1997) and human cells (Armstrong 1998, Wang W 1996). These complexes contain several homologues of proteins identified in the yeast SWI/SNF complex. One of these complexes, human EKLF coactivator remodelling complex 1 (E-RC1), shows functional selectivity for the erythroid specific transcription factor EKLF (Armstrong 1998). E-RC1 and EKLF cooperate in initiating a DNase I hypersensitive region, and both E-RC1 and EKLF are necessary for transcription activation of a β globin promoter *in vitro*. Interestingly a mutant EKLF protein lacking the transactivation domain, required for transcription activation, still shows chromatin remodelling activity and DNase I hypersensitive site formation, indicating that an open chromatin structure can be created prior to the loading of the basal transcription machinery on the DNA. Although E-RC1 seems to function specifically in concert with EKLF the complex could also be purified from HELA cells, suggesting that EKLF is not the only transcription factor that recruits E-RC1 for chromatin remodelling.

Several CRC's, like NURF, CHRAC and ACF have been isolated from *Drosophila*. All three complexes contain the same ATP dependent chromatin remodelling subunit ISWI (Corona 1999). Like SWI/SNF, nucleosome remodelling factor (NURF) mediates binding of transcription factors (Tsukiyama 1995). NURF is involved in the binding of GAGA factor to the *Drosophila* hsp70 promoter. Similar to E-RC1 a hypersensitive site is formed on a chromatin template *in vitro* upon

addition of GAGA, ATP and NURF. ACF also facilitates binding of activators to chromatin packed DNA, and is capable of the assembly of properly spaced nucleosomal arrays (Ito 1997). ACF is a tetrameric complex and co-operates with NAP1 in chromatin assembly during S phase. In contrast to the CRC's mentioned before, a third *Drosophila* complex CHRAC, is not capable of facilitating activators to a promoter (Varga-Weisz 1995, 1997). CHRAC enhances the general accessibility of DNA in chromatin, and functions in chromatin assembly. The complex contains topoisomerase II, although the role of topoisomerase II in this complex remains unclear.

Chromatin remodelling complexes were generally believed to be devoid of histone acetyl transferase activity. However, more recent data show that HDAC's are involved in ATP dependent chromatin remodelling, linking both chromatin modification activities to one complex (Tong 1998, Xue 1998, and Zhang 1998). Two human complexes, NRD and NURD, have been identified yet which are capable of ATP dependent chromatin remodelling as well as histone deacetylation, and purification of chromatin remodelling complexes with histone HAT activity seems to be only a matter of time.

DNase I sensitivity

Areas surrounding active genes or areas surrounding genes with an active history can be distinguished from inactive areas by an increased sensitivity to deoxyribonuclease I (DNase I, Weintraub and Groudine 1976, Groudine 1983). DNase I nicks DNA without the preference for specific sequences and the ability of distinguishing areas with different sensitivities for DNase I provided the first molecular tool for studying chromatin structure *in vivo*. Several other nucleases have been used since, like DNase II and MNase, with identical albeit less clear results. General DNase I sensitivity of chromosomal regions in the chicken β globin locus is 10 fold higher in blood cells compared to the inactive ovalbumin gene (Wood and Felsenfeld 1982). For DNase II and MNase these sensitivities are 6 and 3 fold higher for active compared to inactive regions respectively. Although the exact basis of increased sensitivity remains unclear today, general sensitivity might be explained by the absence of histone H1 and increased histone acetylation in these areas (Smith 1984, Vidali 1978, Hebbes 1994). Specific regions of up to several hundred base pairs usually located within general sensitive areas appear to be hypersensitive to DNase I digestion (Stalder 1980). These hypersensitive sites correspond with binding sites of transcription and transacting factors as well as replication origins and are thought to represent nucleosome free areas as a result of local chromatin remodelling (Wu 1979, Stalder 1980, Scott and Wigmore 1978, Waldeck 1978). Several of these hypersensitive sites were found in the human β globin locus, with erythroid specific and developmentally stable hypersensitive sites located upstream of the area containing the 5 functional genes (Tuan 1985, Grosveld 1987, Forrester 1987). Stage specific hypersensitive sites were found in the promoter areas of the different globin genes (Stalder 1980, Charnay 1984, and Forrester 1986).

Position effect variegation and heterochromatinization

Chromatin can cytologically be separated into euchromatin and heterochromatin. Most active genes reside in euchromatin that is decondensed in interphase and condenses in metaphase whereas heterochromatin contains most inactive genes and remains condensed during the cell cycle. In *Sacheromyces*

cereviciae heterochromatin is found in telomeres and the mating type loci (in *S. pombe* also in centromeres), whereas higher eukaryotes have heterochromatic areas at different locations on the chromosome. Heterochromatin is generally hypoacetylated and comprises high quantities of repetitive sequences. Translocations of euchromatin juxtaposed to heterochromatin show a variable spreading of heterochromatin into the euchromatic area thereby silencing the adjacent genes. Silencing happens in a stochastic and clonal fashion. Some cells express the gene while other do not, with the daughter cell inheriting the expression pattern of the mother. This phenomenon called position effect variegation (PEV) was discovered in *Drosophila.* It was found that when the white gene, necessary for red eye pigmentation, was translocated next to heterochromatin this gene was silenced in part of the cells, resulting in a mosaic coloured eye (Spofford and DeSalle 1991). Studies of transgenic mice with the human β globin and CD2 loci show that deletion of hypersensitive sites in their locus control region (LCR) results in PEV (Milot 1996, Festenstein 1996).

Two non mutually exclusive models explaining PEV have been postulated. The nuclear compartmentalisation model proposes that chromosomal areas reside in specific nuclear territories favouring hetero- or eu-chromatinization, whereas the mass action model explains PEV as a result of different inter and intra cellular concentrations of factors involved in heterochromatin and euchromatin formation.

Evidence for the nuclear compartmentalisation model comes from work in *Drosophila*, which shows that the relative distance of a rearranged white transgene to the juxtaposed heterochromatin determines the number of expressing cells (Dorer 1979). Additional data supporting this model is the trans-inactivation of a wildtype brown allele by a dominant brown allele (bw^D) on the other chromosome (Csink 1996, Dernburg 1996). It was found that the bw^D allele, located in heterochromatin as a consequence of an insertion of a block of heterochromatin in the coding sequence, pairs with the wildtype brown allele in interphase nuclei. This interaction was not found between two wildtype brown alleles. Interestingly the trans-inactivated wildtype brown allele does not show cytological heterochromatinization (Belyaeva 1997), suggesting that the wildtype allele is pulled into a transcriptionally silent area.

Most of the data that supports a mass action model is based on mutations that have been found to result in changes in levels of PEV. These modifiers of PEV can either suppress PEV (Su[var]s) or enhance PEV (E[var]s). Some of these modifiers appear to be the building blocks of heterochromatin, like the *Drosophila* Su(var)2-5/HP1and Su(var)3-7 proteins. Other proteins seem to be involved in replication or chromatin formation. Trithorax-like proteins like GAGA factor, involved in euchromatin formation, have E(var) activity whereas some polycomb group proteins have Su(var) activity (Kassis 1998). Mammalian homologues of PcG and trxG proteins have been cloned like M33, Bmi1 and mbrm, homologues of *Drosophila* Pc, Psc and brahma respectively. Compound transgenics and knockouts of genes coding for these proteins with PEV expression lines of the human β globin locus, show that E(var) and Su(var) activity found in *Drosophila* is conserved in mammals (Milot submitted). Importantly these studies showed that PEV is also modified by the transcription factors EKLF and Sp1. Many polycomb group proteins do not affect levels of PEV when mutated, implying that these proteins have either specific targeting areas, or that only a few proteins involved in Polycomb group protein mediated gene repression are directly involved in heterochromatin formation.

Deletion of hypersensitive sites of the human β globin LCR also revealed another type of position effect. Mouse lines were generated showing expression of the transgene in all red blood cells, albeit at lower levels. Immunofluorescence *in situ* hybridisation studies on sorted cells showed that the β globin gene is only transcribed in certain phases of the cell cycle resulting in a cell timing effect (CTE, Milot 1996).

Elements involved in chromatin domain formation

Several elements with different properties have been implicated in the maintenance and alteration of chromatin structure. Four different elements will be discussed in this paragraph (figure 2). Polycomb and trithorax group proteins bind Polycomb and trithorax responsive elements (PRE, TRE) respectively and play an important role in the establishment and memory of large distance heterochromatic and euchromatic chromatin structures. Insulators are elements that have been implicated in preventing the spreading of heterochromatin into euchromatic regions. Finally locus control regions are dominant tissue specific chromatin opening elements with an additional enhancer function that can act over large distances up to 100 kb away.

Polycomb and trithorax responsive elements

Polycomb group (PcG) proteins were originally identified as mutations resulting in ectopic expression of homeotic genes (Struhl 1981). Several members of the PcG proteins have been cloned and it appears that PcG proteins not only affect homeotic genes but also affect several other genes involved in developmental processes. PcG proteins have been implicated in the establishment and memory of a silent chromatin state (for review see Paro 1995, Pirotta 1998). In contrast to the PcG proteins trithorax group (trxG) proteins are involved in the strong expression of homeotic genes and can at least in part counteract PcG mediated silencing (Cavalli 1998). PcG and trxG proteins act via polycomb responsive elements (PRE) and trithorax responsive elements (TRE) of several hundreds of nucleotides in size (Simon 1993, Orlando 1998). Recent data using the *Drosophila* UB-X locus as a model system show that Pc and Trx localise on the same elements. Localisation appears not to be restricted to PRE's/TRE's but is also found in promoters (Orlando 1998). No specific consensus sequence has been found but the importance of PRE in chromatin memory was clearly shown by a PRE deletion using flp recombinase resulting in a loss of maintenance of the chromatin structure (Busturia 1997). Immunofluorescence studies showing a staining pattern of PcG proteins on *Drosophila* salivary gland chromosomes, suggest that PcG protein would bind to DNA. However the PcG and trxG proteins found to date do not bind DNA, suggesting that additional proteins are involved in DNA binding. Targeting a PcG protein with a DNA binding domain is sufficient to recruit PcG complexes to the DNA and subsequent silencing. Co-localisation and biochemical studies with two different mammalian homologues of the *Drosophila* Pc-G gene Enhancer-of-zeste [E(z)] suggest that different PcG complexes exist *in vivo* (van Lohuizen 1998). Sequence analysis of the initially identified Polycomb protein (Pc) revealed a conserved chromodomain with unknown function. This chromodomain is also found in the modifier of PEV, H1 or Suvar 2-5. Mutations in the H1 chromodomain result in suppression of variegation and show the importance of this domain in heterochromatin formation. Several trxG and PcG members are also modifiers of PEV, indicating that these proteins are involved in heterochromatin formation.

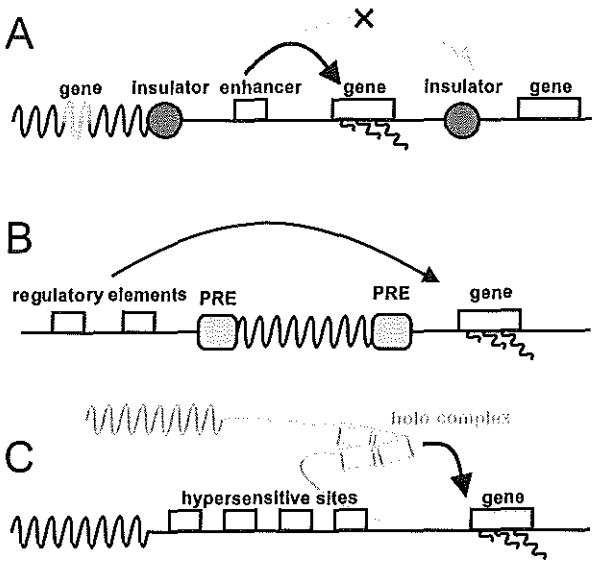


Figure 2. *Models explaining the mechanisms of different elements implicated in chromatin domain formation. (A) Insulator elements prevent enhancer promoter interactions when inserted in between the enhancer and the gene. Insulators also prevent heterochromatin spreading (curled line) into insulated domains. (B) PRE/TRE elements are functional borders of active and inactive domains. Interactions between upstream regulatory elements and the downstream promoter are not blocked by interposed PRE's. (C) LCR's are dominant chromatin opening elements that interact with downstream genes as a holo-complex (grey line). Individual genes compete for LCR activation.*

Interestingly not all PcG proteins tested were modifiers of PEV suggesting either different mechanisms of PcG mediated gene silencing, or different PcG proteins or complexes affecting different loci.

PcG complexes can affect enhancers and promoters over distances up to 30 kb, although the silencing mechanism over these distances is not yet understood. A heterochromatin spreading and a looping model have been postulated to explain the long distance silencing capabilities (Pirotta 1998). Chromatin cross-linking experiments only reveals PcG proteins in the vicinity of PRE's arguing against a spreading model with the PRE as nucleation site for silencing and suggest a looping model with one or multiple PRE's and proto PRE's acting together by looping out intervening DNA (Orlando and Paro 1993). The looping mechanism could also explain the decreasing output of transgenes with increasing copynumber (Fraser

1990), since weak PRE's could now act *in cis* forming a big PRE complex with strong suppressive effects.

Genetic studies of the *Drosophila* BX-C locus show that polycomb mediated silencing affect regulatory segments but can not insulate upstream elements from activating downstream promoters. The BX-C locus consists of three genes, Ultrabithorax, abdominal-A and Abdominal-B that are differentially expressed in different parasegments of the developing fly. Regulation of Abdominal-B expression in parasegments 10-14 is mediated by the *cis* regulatory segments iab5 through iab9, with different iab's being active in different parasegments. Two important domain boundaries have been described so far, MCP situated between iab4 and iab5 and Fab7 separating iab6 and iab7. During early phases of embryogenesis activity of *cis* regulatory elements is initially set by gap and pair-rule proteins, which are only transiently present at this stage of development. The activity state is subsequently maintained by PcG and trxG proteins for which binding sites have been found in the Fab7 and MCP boundary elements (for review see Mihaly 1998). Interestingly Abdominal B expression in proximal parasegments is mediated by proximal iab regulatory segments acting on the distal Abdominal-B gene without interference of the silent iab's in-between. This clearly distinguishes PRE elements from insulators (figure 2).

Insulators

Insulators have been implicated in preventing interactions of regulatory elements of different domains *in cis*, establishing a functional boundary between two adjacent domains. The best-studied insulators are the *Drosophila* scs elements (Kellum and Schedl 1992), the *Drosophila* suppressor of Hairy wing (su[HW]) DNA binding sites (Geyer 1988) and the chicken β globin hypersensitive site 4 (HS4, Chung 1993). Studies with all three insulators show that enhancer activity is abolished when an insulator is inserted between an enhancer and promoter. No enhancer blocking activity is found when the insulator is placed proximal or distal of both the enhancer and promoter (Kellum and Schedl 1992). Insertion of an enhancer and promoter between two insulators protects the transgene from neighbouring position effects (Kellum and Schedl 1991). Insulators don't possess enhancer activity and studies with scs elements and binding sites of the su[Hw] do not reveal changes in chromatin structure leaving the enhancer of an insulated gene accessible for a adjacent non insulated promoter (Cai and Levine 1995, Scott and Geyer 1995). This in contrast with the chicken β globin HS4 that appears to be the border of DNase I sensitivity as well as general histone acetylation (Hebbes 1994). It is however not clear if these differences are the consequence of additional features of the chicken HS4, when compared to the scs or su[Hw] elements. Two different proteins, BEAF-32A and SBP (Udvardy 1999), have been identified in binding the scs hypersensitive site regions. *In situ* hybridisation with antibodies against BEAF32A show staining of several hundreds of interband regions in *Drosophila* polythene chromosomes, indicating that BEAF32A mediated insulation could be a general mechanism of boundary formation in *Drosophila*.

Different models have been proposed to explain the insulator mechanism (Udvardy 1999). In a tracking model transcription factors that track along the DNA from an enhancer towards a promoter are blocked by an insulator inserted in between. A three dimensional model explains the insulating activity by the formation of a specific chromatin loop structure resulting in active domains flanked by

insulators making spherical interactions between different domains impossible. In the transcriptional decoy model insulators compete with the promoter by attracting enhancer activity. Reports using two interlocked plasmids (catenanes), one of which containing a promoter the other an enhancer, clearly show enhancer activity on a promoter *in trans* (Herendeen 1992, Krebs and Dunaway 1998). Enhancer activity was abrogated by either placing the gene or the enhancer between scs elements (Krebs and Dunaway 1998). This observation strongly argues against a tracking model with an insulator block *in cis*. It is likewise difficult to envisage how insulators trap enhancer activity only when placed in between an enhancer and a promoter and not upstream of the enhancer, making the domain looping model the favoured model for explaining insulator function.

Locus control regions

The β globin locus is among the best-studied eukaryotic multi gene loci. Initial transgenic studies with globin genes integrated in the mouse genome made clear that these sequences were not sufficient for proper expression of the transgene. The same observation was made in a Dutch thalassemia patient (van der Ploeg 1980), showing an absence in globin mRNA despite the presence of intact adult globin genes. The patient showed a deletion extending into the proximal region of the locus, resulting in an inactive chromatin configuration of the remaining sequences of the mutated locus (Kioussis 1983). Five developmentally stable and erythroid specific hypersensitive sites were identified in the region upstream of the area containing the five functional genes. Transgenic studies with this region, designated locus control region (LCR), linked to a globin gene revealed copy number dependent and position independent expression (Grosveld 1987).

Several other LCR's have been identified since showing correct chromatin activation and developmental expression of linked transgene (Lang 1989, Bonifer 1994). It was found that the different globin genes compete for activation by the LCR during development (Choi and Engel 1988, Enver 1990, Hanscombe 1991), and that competition is dependent on the relative distance from the LCR and the promoter strength of active genes (Dillon 1997). Deletions of individual hypersensitive sites of the human β globin LCR do not show a profound effect on individual genes during development, but results in a reduction in expression of all different genes during development (Bungert 1995, Milot 1996, Peterson 1996). These data suggest that the different hypersensitive sites act as a holo-complex in order to activate distal genes. Interestingly *in situ* hybridisation studies on mouse lines with deletions of hypersensitive sites of the LCR reveal that the transgene is now susceptible to position effect variegation, when integrated into heterochromatic regions of the mouse genome (Milot 1996). The same result is found for deletions of hypersensitive sites in the LCR of the human CD2 locus (Festenstein 1996). Both loci show PEV when integrated in centromeric regions, whereas centromeric integration of either locus with a full LCR shows normal expression, indicating that LCR's are important instruments in overcoming heterochromatinization. LCR's have also been implicated in histone acetylation as has been found for the chicken β globin locus. A correlation between general histone acetylation and general DNase I sensitivity was established (Hebbes 1994). Two recent publications suggest that general DNase I sensitivity is not a direct effect of the LCR acting on chromatin, since a deletion of the human and mouse LCR does not result in abrogation of general sensitivity and hypersensitive site formation in the locus (Reik 1998, Epner 1998). Although the data show that

transcription activation requires the LCR, it remains questionable if the LCR is required for chromatin activation (Epner 1998). The transgenic mice that have been produced with LCR constructs integrated in the mouse genome show that the LCR is sufficient to activate the globin locus. There may however be epigenetic information and/or DNA sequences left in the areas not deleted by Epner *et al.* (1998), that were sufficient to activate the locus.

References

Allan J., Mitchell T., Harborne N., Bohm L., Crane-Robinson C. 1986. Roles of H1 domains in determining higher order chromatin structure and H1 location. *J Mol Biol* **187**:591-601.

Almer A., Rudolph H., Hinnen A., Horz W. 1986. Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. *EMBO J* **5**:2689-96.

Armstrong J.A., Bieker J.J., Emerson B.M. 1998. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* **95**:93-104.

Ashe H.L., Monks J., Wijgerde M., Fraser P., Proudfoot N.J. 1997. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11**:2494-509.

Banerji J., Rusconi S., Schaffner W. 1981 Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**:299-308.

Belyaeva E.S., Koryakov D.E., Pokholkova G.V., Demakova O.V., Zhimulev I.F. 1997. Cytological study of the brown dominant position effect. *Chromosoma* **106**:124-32.

Bentley D.L., Groudine M. 1986. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. *Nature* **321**:702-6.

Birse C.E., Minvielle-Sebastia L., Lee B.A., Keller W., Proudfoot N.J. 1998. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280**:298-301.

Bonifer C., Yannoutsos N., Kruger G., Grosveld F., Sippel A.E. 1994. Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. *Nucleic Acids Res* **22**:4202-10.

Boyes J., Byfield P., Nakatani Y., Ogryzko V. 1998. Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* **396**:594-8.

Brown J.L., Mucci D., Whiteley M., Dirksen M.L., Kassis J.A. 1998. The *Drosophila* Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell* **1**:1057-64

Brownell J.E., Zhou J., Ranalli T., Kobayashi R., Edmondson D.G., Roth S.Y., Allis C.D. 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**:843-51.

Bungert J., Dave U., Lim K.C., Lieuw K.H., Shavit J.A., Liu Q., Engel J.D. 1995. Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev* **9**:3083-96.

Busturia A., Wightman C.D., Sakonju S. 1997. A silencer is required for maintenance of transcriptional repression throughout *Drosophila* development. *Development* **124**:4343-50.

Cai H., Levine M. 1995 Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. *Nature* **376**:533-6.

Cairns B.R., Lorch Y., Li Y., Zhang M., Lacomis L., Erdjument-Bromage H., Tempst P., Du J., Laurent B., Kornberg R.D. 1996. RSC, an essential, abundant chromatin-remodeling complex. *Cell* **87**:1249-60.

Cavalli G., Paro R. 1998. The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**:505-18.

Charnay P., Treisman R., Mellon P., Chao M., Axel R., Maniatis T. 1984. Differences in human alpha- and beta-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. *Cell* **38**:251-63.

Cho E.J., Takagi T., Moore C.R., Buratowski S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11**:3319-26.

Cho H., Orphanides G., Sun X., Yang X.J., Ogryzko V., Lees E., Nakatani Y., Reinberg D. 1998. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol* **18**:5355-63.
Chodosh L.A., Fire A., Samuels M., Sharp P.A. 1989. 5,6-Dichloro-1-beta-D-ribofuranosyl-benzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *J Biol Chem* **264**:2250-7.
Choi O.R., Engel J.D. 1988. Developmental regulation of beta-globin gene switching. *Cell* **55**:17-26.
Chung J.H., Whiteley M., Felsenfeld G. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **4**:505-14.
Clarke D.J., O'Neill L.P., Turner B.M. 1993. Selective use of H4 acetylation sites in the yeast Saccharomyces cerevisiae. *Biochem J* **294**:557-61.
Corden J.L., Patturajan M. 1997. A CTD function linking transcription to splicing. *Trends Biochem Sci* **22**:413-6.
Corona D.F., Langst G., Clapier C.R., Bonte E.J., Ferrari S., Tamkun J.W., Becker P.B. 1999. ISWI is an ATP-dependent nucleosome remodeling factor. *Mol Cell* **3**:239-45.
Krebs J.E., Dunaway M. 1998. The scs and scs' insulator elements impart a cis requirement on enhancer-promoter interactions. *Mol Cell* **1**:301-8.
Csink A.K., Henikoff S. 1996. Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature* **381**:529-31.
Dantonel J.C., Murthy K.G., Manley J.L., Tora L. 1997. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* **389**:399-402.
Dernburg A.F., Broman K.W., Fung J.C., Marshall W.F., Philips J., Agard D.A., Sedat J.W. 1996. Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell* **85**:745-59.
Dillon N., Trimborn T., Strouboulis J., Fraser P., Grosveld F. 1997. The effect of distance on long-range chromatin interactions. *Mol Cell* **1**:131-9.
Dorer D.R., Henikoff S. 1979. Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics* **147**:1181-90.
Enver T., Raich N., Ebens A.J., Papayannopoulou T., Costantini F., Stamatoyannopoulos G. 1990. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* **344**:309-13.
Epner E., Reik A., Cimbora D., Telling A., Bender M.A., Fiering S., Enver T., Martin D.I., Kennedy M., Keller G., Groudine M. 1998. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell* **2**:447-55.
Festenstein R., Tolaini M., Corbella P., Mamalaki C., Parrington J., Fox M, Miliou A, Jones M., Kioussis D.M. 1996. Locus control region function and heterochromatin-induced position effect variegation. Science **271**:1123-5.
Forrester W.C., Thompson C., Elder J.T., Groudine M. 1986. A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A* **83**:1359-63.
Forrester W.C. Takegawa S., Papayannopoulou T., Stamatoyannopoulos G., Groudine M. 1987. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res* **15**:10159-77.
Fraser P., Hurst J., Collis P., Grosveld F. 1990. DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res* **18**:3503-8.

Georgakopoulos T., Thireos G. 1992. Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. *EMBO J* **11**:4145-52.
Geyer P.K., Green M.M., Corces V.G. 1988. Mutant gene phenotypes mediated by a *Drosophila* melanogaster retrotransposon require sequences homologous to mammalian enhancers. *Proc Natl Acad Sci U S A* **85**:8593-7.
Grant P.A., Duggan L., Cote J., Roberts S.M., Brownell J.E., Candau R., Ohba R., Owen-Hughes T., Allis C.D., Winston F., Berger S.L., Workman J.L. 1997. Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev* **11**:1640-50.
Grosveld F., van Assendelft G.B., Greaves D.R., Kollias G. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**:975-85.
Groudine M., Kohwi-Shigematsu T., Gelinas R,. Stamatoyannopoulos G., Papayannopoulou T. 1983. Human fetal to adult hemoglobin switching: changes in chromatin structure of the beta-globin gene locus. *Proc Natl Acad Sci U S A* **80**:7551-5.
Hanscombe O., Whyatt D., Fraser P., Yannoutsos N., Greaves D., Dillon N., Grosveld F. 1991. Importance of globin gene order for correct developmental expression. *Genes Dev* **5**:1387-94.
Hassig C.A., Fleischer T.C., Billin A.N., Schreiber S.L., Ayer D.E. 1997. Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell* **89**:341-7.
Hebbes T.R., Clayton A.L., Thorne A.W., Crane-Robinson C. 1994. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J* **13**:1823-30.
Henikoff S. 1997. Nuclear organization and gene expression: homologous pairing and long-range interactions. *Curr Opin Cell Biol* **9**:388-95.
Herendeen D.R., Kassavetis G.A., Geiduschek E.P. 1992. A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science* **256**:1298-303.
Imbalzano A.N., Kwon H., Green M.R., Kingston R.E. 1994. Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* **370**:481-5.
Imhof A., Yang X.J., Ogryzko V.V., Nakatani Y., Wolffe A.P., Ge H. 1997. Acetylation of general transcription factors by histone acetyltransferases. *Curr Biol* **7**:689-92.
Ito T., Bulger M., Pazin M.J., Kobayashi R., Kadonaga J.T. 1997. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* **90**:145-55.
Izban M.G., Luse D.S. 1992. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem* **267**:13647-55.
Jackson V., Shires A., Tanphaichitr N., Chalkley R. 1976. Modifications to histones immediately after synthesis. *J Mol Biol* **104**:471-83.
Jenuwein T., Forrester W.C., Fernandez-Herrero L.A., Laible G., Dull M., Grosschedl R.1997. Extension of chromatin accessibility by nuclear matrix attachment regions. *Nature* **385**:269-72.
Jeppesen P., 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**:281-9.
Kadonaga J.T. 1998. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* **92**:307-13.

Kadosh D., Struhl K. 1997. Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell* **89**:365-71.
Kadosh D., Struhl K. 1998a. Histone deacetylase activity of Rpd3 is important for transcriptional repression in vivo. *Genes Dev* **12**:797-805.
Kadosh D., Struhl K. 1998b. Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo. *Mol Cell Biol* **18**:5121-7.
Kassis J.A., Brock H.W. 1998. Enhancer of Polycomb is a suppressor of position-effect variegation in *Drosophila* melanogaster. *Genetics* **148**:211-20.
Keaveney M., Struhl K. 1998. Activator-mediated recruitment of the RNA polymerase II machinery is the predominant mechanism for transcriptional activation in yeast. *Mol Cell* **1**:917-24.
Kellum R., Schedl P. 1991. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**:941-50.
Kellum R., Schedl P. 1992. A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol* **12**:2424-31.
Kioussis D., Vanin E., deLange T., Flavell R.A., Grosveld F.G. 1983. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* **306**:662-6.
Kioussis D. 1996. Locus control region function and heterochromatin-induced position effect variegation. *Science* **271**:1123-5.
Koleske A.J., Young R.A. 1994. An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**:466-9.
Kuo M.H., Brownell J.E., Sobel R.E., Ranalli T.A., Cook R.G., Edmondson D.G., Roth S.Y., Allis C.D. 1996. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature* **383**:269-72.
Lamond A.I., Earnshaw W.C. 1998. Structure and function in the nucleus. *Science* **280**:547-53.
Lang G., Wotton D., Owen M.J., Sewell W.A., Brown M.H., Mason D.Y., Crumpton M.J., Kioussis D. 1988. The structure of the human CD2 gene and its expression in transgenic mice. *EMBO J* **7**:1675-82.
Laurent B.C., Treitel M.A., Carlson M. 1990. The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and proline-rich transcriptional activator that affects expression of a broad spectrum of genes. *Mol Cell Biol* **10**:5616-25.
Lee D.Y., Hayes J.J., Pruss D., Wolffe A.P. 1993. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72**:73-84.
Lorch Y., LaPointe J.W., Kornberg R.D. 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**:203-10.
Lucchini R., Sogo J.M. 1995. Replication of transcriptionally active chromatin. *Nature* **374**:276-80.
Marshall N.F., Peng J., Xie Z., Price D.H. 1996 Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271**:27176-83.
Martin D.I., Fiering S., Groudine M. 1996. Regulation of beta-globin gene expression: straightening out the locus. *Curr Opin Genet Dev* **6**:488-95.
McCracken S., Fong N., Yankulov K., Ballantyne S., Pan G., Greenblatt J., Patterson S.D., Wickens M., Bentley D.L. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**:357-61.

McKnight S.L. 1996. Transcription revisited: a commentary on the 1995 Cold Spring Harbor Laboratory meeting, "Mechanisms of Eukaryotic Transcription". *Genes Dev* **10**:367-81.

Mihaly J., Hogga I., Barges S., Galloni M., Mishra R.K., Hagstrom K., Muller M., Schedl P., Sipos L., Gausz J., Gyurkovics H., Karch F. 1998. Chromatin domain boundaries in the Bithorax complex. *Cell Mol Life Sci* **54**:60-70.

Milot E., Strouboulis J., Trimborn T., Wijgerde M., de Boer E., Langeveld A., Tan-Un K., Vergeer W., Yannoutsos N., Grosveld F., Fraser P. 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87**:105-14.

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., Allis C.D.1996. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**:1261-70.

Moreau P., Hen R., Wasylyk B., Everett R., Gaub M.P., Chambon P. 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* **9**:6047-68.

Mueller-Storm H.P., Sogo J.M., Schaffner W. 1989. An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell* **58**:767-77.

Ogryzko V.V., Schiltz R.L., Russanova V., Howard B.H., Nakatani Y. 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**:953-9.

O'Neill L.P., Turner B.M. 1995. Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *EMBO J* **14**:3946-57.

Orlando V., Paro R. 1993. Mapping Polycomb-repressed domains in the bithorax complex using *in vivo* formaldehyde cross-linked chromatin. *Cell* **75**:1187-98.

Orlando V., Jane E.P., Chinwalla V., Harte P.J., Paro R. 1998. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J* **17**:5141-50.

Ossipow V., Tassan J.P., Nigg E.A., Schibler U. 1995. A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell* **83**:137-46.

Paro R. 1995. Propagating memory of transcriptional states. *Trends Genet* **11**:295-7.

Peng J., Zhu Y., Milton J.T., Price D.H. 1998. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev* **12**:755-62.

Peterson K.R., Clegg C.H., Navas P.A., Norton E.J., Kimbrough T.G., Stamatoyannopoulos G. 1996. Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A* **93**:6605-9.

Pirrotta V. 1998. Polycombing the genome: PcG, trxG, and chromatin silencing. *Cell* **93**:333-6.

Ptashne M. 1988. How eukaryotic transcriptional activators work. *Nature* **335**:683-9.

Recht J., Osley M.A. 1999. Mutations in both the structured domain and N-terminus of histone H2B bypass the requirement for swi-Snf in yeast. *EMBO J* **18**:229-40.

Reeder R.H., Lang W.H. 1997 Terminating transcription in eukaryotes: lessons learned from RNA polymerase I. *Trends Biochem Sci* **22**:473-7.

Reik A., Telling A., Zitnik G., Cimbora D., Epner E., Groudine M. 1998. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol* **18**:5992-6000.

Scott W.A., Wigmore D.J. 1978. Sites in simian virus 40 chromatin which are preferentially cleaved by endonucleases. *Cell* **15**:1511-8.
Scott K.S., Geyer P.K. 1995. Effects of the su(Hw) insulator protein on the expression of the divergently transcribed *Drosophila* yolk protein genes. *EMBO J* **14**:6258-67.
Shermoen A.W., O'Farrell P.H. 1991. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* **67**:303-10.
Shilatifard A., Conaway J.W., Conaway R.C. 1997. Mechanism and regulation of transcriptional elongation and termination by RNA polymerase II. *Curr Opin Genet Dev* **7**:199-204.
Simon J., Chiang A., Bender W., Shimell M.J., O'Connor M. 1993. Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Dev Biol* **158**:131-44.
Sinclair D.A., Clegg N.J., Antonchuk J., Milne T.A., Stankunas K., Ruse C., Grigliatti T.A., Smale S.T., Baltimore D. 1989. The "initiator" as a transcription control element. *Cell* **57** : 103-13.
Smith R.D., Yu J., Seale R.L. 1984. Chromatin structure of the beta-globin gene family in murine erythroleukemia cells. *Biochemistry* **23**:785-90.
Sogo J.M., Stahl H., Koller T., Knippers R. 1986. Structure of replicating simian virus 40 minichromosomes. The replication fork, core histone segregation and terminal structures. *J Mol Biol* **189**:189-204.
Spofford J.B., DeSalle R. 1991. Nucleolus organizer-suppressed position-effect variegation in *Drosophila* melanogaster. *Genet Res* **57**:245-55.
Stalder J., Larsen A., Engel J.D., Dolan M., Groudine M., Weintraub H. 1980. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* **20**:451-60.
Stern M., Jensen R., Herskowitz I. 1984. Five SWI genes are required for expression of the HO gene in yeast. *J Mol Biol* **178**:853-68.
Struhl G.A. 1981. homoeotic mutation transforming leg to antenna in *Drosophila*. *Nature* **292**:635-8.
Struhl K. 1996. Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell* **84**:179-82.
Su W., Jackson S., Tjian R., Echols H. 1991. Looping between sites for transcriptional activation: self-association of DNA-bound Sp1. *Genes Dev* **5**:820-6.
Tong J.K., Hassig C.A., Schnitzler G.R., Kingston R.E., Schreiber S.L. 1998. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* **395**:917-21.
Tsukiyama T., Wu C.1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**:1011-20.
Tuan D., Solomon W., Li Q., London I.M 1985. The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A* **82**:6384-8.
Turner B.M., Birley A.J., Lavender J. 1992. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell* **69**:375-84.
Udvardy A. 1999 Dividing the empire: boundary chromatin elements delimit the territory of enhancers. *EMBO J* **18**:1-8.
Van der Ploeg L.H., Konings A., Oort M., Roos D., Bernini L., Flavell R.A. 1980. gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature* **283**:637-42.

van Lohuizen M., Tijms M., Voncken J.W., Schumacher A., Magnuson T., Wientjens E. 1998. Interaction of mouse polycomb-group (Pc-G) proteins Enx1 and Enx2 with Eed: indication for separate Pc-G complexes. *Mol Cell Biol* **18**:3572-9.

Varga-Weisz P.D., Blank T.A., Becker P.B. 1995. Energy-dependent chromatin accessibility and nucleosome mobility in a cell-free system. *EMBO J* **14**:2209-16.

Varga-Weisz P.D., Wilm M., Bonte E., Dumas K., Mann M., Becker P.B. 1997. Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* **388**:598-602.

Verreault A., Kaufman P.D., Kobayashi R., Stillman B. 1996. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**:95-104

Vettese-Dadey M., Grant P.A., Hebbes T.R., Crane-Robinson C., Allis C.D., Workman J.L. 1996. Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *EMBO J* **15**:2508-18.

Vidali G., Boffa L.C., Bradbury E.M., Allfrey V.G. 1978. Butyrate suppression of histone deacetylation leads to accumulation of multiacetylated forms of histones H3 and H4 and increased DNase I sensitivity of the associated DNA sequences. *Proc Natl Acad Sci U S A* **75**:2239-43.

Waldeck W., Fohring B., Chowdhury K., Gruss P., Sauer G. 1978. Origin of DNA replication in papovavirus chromatin is recognized by endogenous endonuclease. *Proc Natl Acad Sci U S A* **75**:5964-5968.

Wang W., Xue Y., Zhou S., Kuo A., Cairns B.R., Crabtree G.R. 1996. Diversity and specialization of mammalian SWI/SNF complexes. *Genes Dev* **10**:2117-30.

Weintraub H., Groudine M. 1976. Chromosomal subunits in active genes have an altered conformation. *Science* **193**:848-56.

Wilson C.J., Chao D.M., Imbalzano A.N., Schnitzler G.R., Kingston R.E., Young R.A. 1996. RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* **84**:235-44.

Winston F., Carlson M. 1992. Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet* **8**:387-91.

Wood W.I., Felsenfeld G. 1982. Chromatin structure of the chicken beta-globin gene region. Sensitivity to DNase I, micrococcal nuclease, and DNase II. *J Biol Chem* **257**:7730-6.

Wu C., Wong Y.C., Elgin S.C. 1979. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* **16**:807-14.

Xue Y., Wong J., Moreno G.T., Young M.K., Cote J., Wang W. 1998. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* **2**:851-61.

Yang W.M., Inouye C., Zeng Y., Bearss D., Seto E. 1996. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. *Proc Natl Acad Sci U S A* **93**:12845-50.

Yokota H., van den Engh G., Hearst J.E., Sachs R.K., Trask B.J. 1995. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J Cell Biol* **130**:1239-49.

Zhang W., Bieker J.J. 1998. Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. *Proc Natl Acad Sci U S A* **95**:9855-60.

Zhang Y., LeRoy G., Seelig H.P., Lane W.S., Reinberg D. 1998. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**:279-89.

Zawel L., Kumar K.P., Reinberg D. 1995. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev* 9:1479-90.

Chapter 2

Erythropoiesis and the regulation of globin gene expression.

An introduction

Parts of this chapter have been published in: Current Opinion in Hematology 5: 139-144. 1998. Mechanisms of developmental regulation in globin loci. P. Fraser, J. Gribnau and T. Trimborn.

Erythropoiesis

Millions of blood cells are produced every minute in order to keep up with the needs of the human defence system and oxygen transport through the body. The process of self-renewal and differentiation of hematopoietic stem cells (HSC) into the different blood lineages is called hematopoiesis. During hematopoiesis, differentiating cells lose their proliferating capacity and gain the lineage specific attributes. HSC are pluripotent cells that can give rise to many different cell types of the lymphoid, myeloid and erythroid lineages, and different sites of hematopoiesis are used during development. In mammals primitive hematopoiesis resides in blood islands of the yolk sac and gives rise to nucleated blood cells in the embryo. In the fetus hematopoiesis moves to the fetal liver after which the bone marrow becomes the major site of hematopoiesis shortly before birth and thereafter.

Human red blood cells comprise up to 50% of the total blood volume, every red cell being replaced on average 120 days after entering the circulation, thus enormous amounts of red blood cells have to be produced every day. The differentiation process from hematopoetic stem cell till red blood cell is called erythropoiesis. It takes place via a series of intermediate precursor cells, like colony and burst forming units (CFU-E and BFU-E), that lose proliferative capacity upon differentiation (Figure 1). Meanwhile erythroid specific genes are activated during differentiation eventually leading to a mature red cell. Primitive embryonic erythropoiesis gives rise to nucleated blood cells whereas definitive fetal and adult erythropoiesis result in an enucleated red cell.

Red cells are specialised in oxygen transport through the body. In mammals the vast majority of the soluble protein of a red blood cell comprises the oxygen

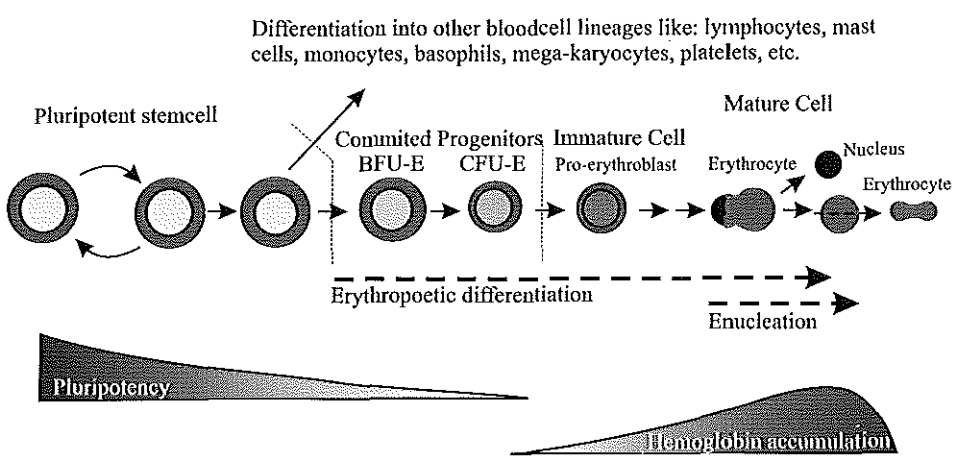


Figure 1. *Schematic representation of human erythropoeisis. Red cells arise from pluripotent stem cells via intermediate precursor steps and lose their proliferative capacity upon differentiation. The nuclear size decreases during differentiation as a consequence of increasing chromatin compaction. In the meantime red cell specific trans-acting factors accumulate resulting in an increase of hemoglobin. Primitive erythropoiesis gives rise to a nucleated red cell whereas definitive erythropoiesis results in an enucleated blood cell.*

carrier protein hemoglobin. Hemoglobin is a hetero-tetrameric protein that consists of two α-like and two β-like globin chains with one heme group attached to each individual globin chain. Most mammals have different α-like and β-like globins expressed during development in order to get hemoglobin with different affinities for oxygen. The affinity for oxygen early during development is higher than the adult hemoglobin, which makes it easier for the embryo to obtain oxygen from the mother. Most of the oxygen transport through the body is mediated by hemoglobin although a monomeric myoglobin is found in muscles. Sequence comparison of α-like, β-like and myoglobin suggest that these globins diverged from the same ancestor about 450 million years ago (Hardison 1996). Oxygen carriers have even been found in several plants species that use leghemoglobin to provide oxygen to symbiotic bacteria that need oxygen for the reduction of nitrogen (Powell 1988).

Hemoglobinopathies

More than 250 million people are carriers of hereditary blood disorders (hemoglobinopathies). Most of these disorders are the result of mutations or deletions of genic or regulatory elements of α-like or β-like globin genes and have been very important for our current understanding of globin gene regulation. The three most common disorders, namely thalassemia, sickle cell anemia and HPFH will be discussed below.

Thalassemias

Thalassemic patients have a disturbed balance in the α/β globin ratio and indicate the importance of co-ordinated expression of α- and β-like globin. Severe disturbance of this α/β globin ratio, as a consequence of reduced expression of α-like or β-like globin, leads to microcytic and hypochromatic cells that can give rise to anemia. Both α and β thalassemias can be distinguished in deletion type and non-deletion type thalassemia. The deletion type thalassemia's usually have large deletions encompassing genic and/or regulatory regions whereas the non-deletion type thalassemia is the result of base mutations in the promoter area or within the gene. The latter giving rise to frameshifts, nonsense mutations, RNA processing mutants or a decrease in transcription. Many deletion and non-deletion type β thalassemia mutations have been described (Collins 1984, Stamatoyanopoulos 1996), but only a few of them played a role in the discovery of important regulatory areas in the globin loci. For example the Dutch deletion type β thalassemic patient played an important role in the discovery of the locus control region. This patient has an intact adult β globin promoter and gene sequences but the gene is not transcribed (van der Ploeg 1980, Kioussis 1983, Taramelli 1986). Deletions upstream of the β globin gene (figure 2), encompassing important regulatory elements, appeared to be the cause of the absence of transcription (Grosveld 1987). Non-deletion type β thalassemias marked the importance of promoter elements like the β globin CACC box. Mutations in this sequence result in a severe down regulation of the β globin gene expression (Kulozik 1991), and it was later shown that this CACC box is the binding site for the β globin specific transcription factor EKLF. Interestingly thalassemia patients have been reported with intact globin loci, possibly having mutations or deletions of genes coding for trans-acting factors that affect globin gene expression. A good example is a recently described mutation in the SWI/SNF related XH2 gene, resulting in α thalassemia which strongly suggests that this gene directly regulates the α globin genes (Gibbons 1995).

Sickle Cell Anemia

Sickle cell anemia is caused by a single T→A base pair mutation in the coding region of the β globin gene resulting in the substitution of a valine for glutamic acid of the 6th codon. The resulting HbS is soluble in oxygenated form but forms large polymers and precipitates upon deoxygenation (for review see Stamatoyanopoulos 1996). Precipitation can be prevented by relative low levels of wildtype adult or fetal hemoglobin and is not observed in sickle trait individuals (AS) with only a single mutated allele. In most cases, patients with both alleles mutated (SS) have red cells with severe morphological changes that appear sickle shaped, as a consequence of precipitating HbS. Repeated oxy- and deoxy-genation leads to ireversibly sickled cells (ISC) which have been considered to be the major pathophysiological cause of sickle cell disease. Sickle cell patients have modest to severe anemia and crisis of peripheral tissues caused by vessel occlusion of trapped sickled cells. Interestingly the frequency of sickle cell anemia (and thalassemia) geographically overlaps the incidence of malaria caused by *Plasmodium falciparum*. Sickle trait individuals appear to have a selective resistance to malaria compared to normal individuals. This advantage also explains the high frequency of the sickle allele, despite the mortality of SS patients.

HPFH

Hereditary persistence of fetal hemoglobin (HPFH) is considered a disorder rather than a disease, since the persistence of fetal hemoglobin does not have clinical consequences for individuals. However most of the HPFH deletions and mutations have been found in combination with sickle cell anemia or β thalassemia. Increased fetal hemoglobin (HbF) levels reduces disease severity of patients with sickle cell anemia by increasing the solubility of HbS. Different point mutations have been reported in the promoter areas of both the $^{G}\gamma$ and $^{A}\gamma$ genes resulting in increased

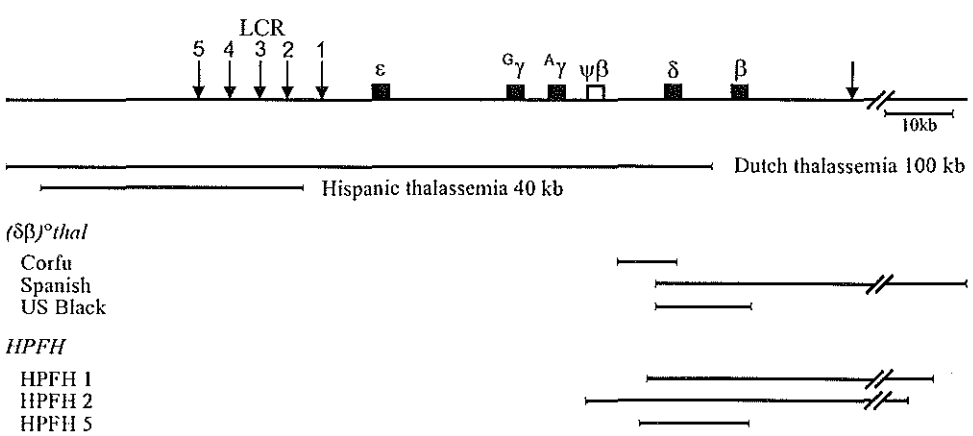


Figure 2. *Localization and size of different deletions resulting in (δβ)°thalessemia and heriditary persistance of fetal hemoglobin (HPFH) phenenotypes. The figure represents the different mutations that will be discussed in this chapter.*

levels of HbF. HbF expression is pancellular and levels vary between different mutations but can be elevated up to 30%. Most of these point mutations disrupt important sequences of the $^{G}\gamma$ and $^{A}\gamma$ promoters, like a CCAAT box at -117 bp, a GATA binding site at -175 bp and a C-rich sequence at -202 bp (see figure 4). Several deletions giving rise to HPFH have also been reported. These deletions usually encompass the adult β globin gene bringing distal hypersensitive sites in close proximity to the γ genes. Comparison of such deletions with deletions that give rise to β thalassemia reveals a very small area between the ψβ and δ genes (Camaschella 1990, Anagnou 1985, Palena 1994). The differences in phenotypes could be the result of downstream sequences that are brought in, but could also be the consequence of different 5' breakpoints. Considering the last possibility the US Black 6 thalassemia and HPFH 5 deletions delineate a 700 base pair area that marks the minimal difference of the 5' breakpoints between a β thalassemia and HPFH phenotype. In the US Black 6 thalassmia this region is still present but is absent in the HPFH 5 deletion, suggesting that the area is important silencing of the γ genes (Huisman 1974, Mears 1978, Ottolenghi 1982, Tuan 1983). However, the Corfu deletion that includes this area does not show increased levels of γ expression (Kulozik 1988), and recent data using transgenic mice with this specific area deleted show that the area is important for the activation of the adult β globin gene but is not required for silencing of the γ genes (Calzolari 1999). These data also suggest that the areas that are brought in close proximity to the γ genes are responsible for elevated levels of HbF.

Structure of the globin loci

The human β globin locus spans about 70 kb on chromosome 11 and comprises 5 functional erythroid specific genes 5'-ε-$^{G}\gamma$-$^{A}\gamma$-δ-β-3' (reviewed in Collins and Weissman 84). The genes are arranged in the order they are expressed during development. This typical organisation with early expressed genes located 5' relative to late expressed genes is found in many organisms that use hemoglobin as oxygen carrier and is probably the result of duplications and divergence of a common ancestor during evolution (Hardison 1996). The ε gene is expressed early during development when erythropoiesis is located in the blood islands of the yolk sac (figure 3). During erythropoiesis in the fetal liver ε expression is silenced with a concomitant increase of both γ gene products. The second switch from γ to δ and β globin expression takes place around birth when the bone marrow becomes the major source of erythropoiesis. Five erythroid specific developmentally stable hypersensitive sites have been identified upstream of the globin genes (Tuan 1985, Grosveld 1987, Forrester 1987), and β thalassemia patients with a deletion encompassing these hypersensitive sites but leaving the adult globin gene intact indicated the importance of this area for proper β globin expression (van der Ploeg 1980, Kioussis 1983, Taramelli 1986). Transgenic studies finally showed that these hypersensitive sites, designated the locus control region (LCR), were sufficient to drive high level, tissue specific expression of linked transgenes in a copy number dependent and position independent manner (Grosveld 1987).

An similar organisation has been found for the α globin locus which has three different functional erythroid specific genes in the order 5'-ζ-α2-α1-3' (reviewed in Higgs 1993). The expression of the α genes displays only one switch from ζ globin, which is predominantly expressed in embryonic blood cells, to α globin that is expressed in the embryonic blood, the fetal liver and the adult bone marrow. An

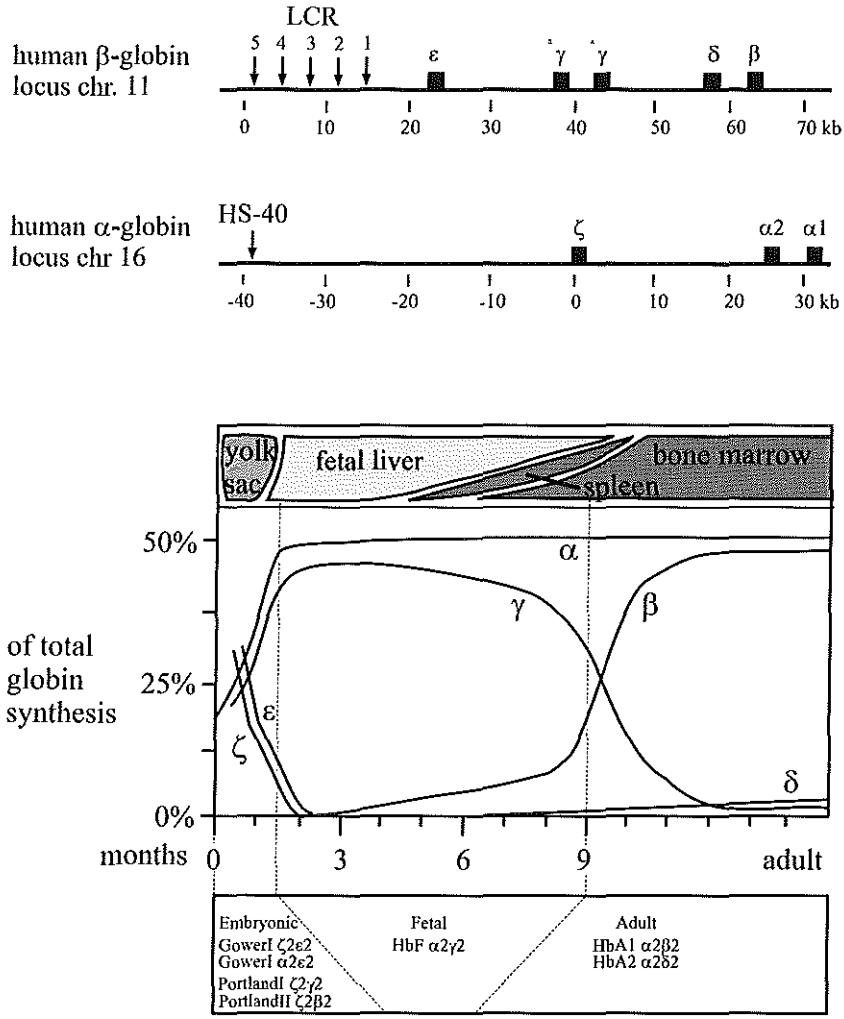


Figure 3. *Organization of the human α and β globin loci, the expression of the globin genes and the site of erythropoiesis during development. The α-like and β-like proteins that are expressed at the same stage of development can form different types of hemoglobin with each hetero-tetramer having specific oxygen affinity characteristics.*

erythroid specific hypersensitive site was found 40 kb upstream of the embryonic ζ gene. Although this hypersensitive site is essential for α globin expression, as shown by knock out studies in cell hybrids, this hypersensitive site does not confer position independent and copy number dependent expression and can therefore not be considered an LCR. The differentially expressed α-like and β-like globins result in a

wide variety of different hemoglobins, each having specific affinities for oxygen. This fine-tuning of the affinity for oxygen during development enables the developing embryo to obtain enough oxygen from the mother (figure 3).

Globin gene structure.

Vertebrates have relative conserved globin gene structure consisting of three exons and two introns. They vary in size from 750 bp for the human α gene up to 1600 bp for the human β globin gene. The human β globin gene exon 1 is 142 bp of which, 92 bp is coding sequence and 50 bp 5' untranslated region (UTR). The second exon is 222 bp in size and exon 3 is 261 bp of which 114 bp coding sequence and a 147 bp 3'UTR. Differences in gene size between different globin genes are mostly the result of varying sizes of the introns. Splicing out of the intron sequences and posttranscriptional addition of a 100 to 200 bp poly-A tail results in a β globin messenger RNA of about 700 bp sequence coding for a 17 kD protein. Although intron sequences are not required for any coding these sequences play an important role in transport of the messenger toward the cytoplasm. *In situ* hybridization studies show that deletion of intron 2 of the human β globin genes results in an accumulation of messenger at the site of transcription (Antoniou 1999). Promoter sequences are, like the gene structure, very well conserved within the different globin genes (Collins and Weissman 1984). The globin promoter contains a TATA box preceded by one or two CCAAT boxes, which are found in many eukaryotic promoters (figure 4). One or two CACC box elements are located upstream of the CCAAT box (Myers 1986).

The ε globin gene promoter has several putative transcription factor binding sites for GATA1 and Sp1 proteins. The promoter contains a single CACC, CCAAT and TATA box. Studies with transgenic mice show that the ε gene requires the LCR for activation and is autonomously silenced during development (Shih 1990, Raich 1990). Silencing is dependent on a proximal silencer element that contains GATA1 and YY1 binding sites. Interestingly deletion of this silencer element, in the context of the whole locus, results in complete absence of epsilon transcription indicating a dual role for this element in activation and silencing (Liu 1997).

The promoters of the two γ globin genes have been well characterised due to the medical relevance of reactivation of the γ genes in adults. Patient data show that a slight elevation of fetal hemoglobin reduces disease severity of sickle cell anemia and β thalassemia indicating the therapeutic potential of factors that could re-establish γ gene expression in the adult. Both γ globin promoters have a TATA box, two CCAAT boxes and one CACC box (Collins and Weissman 1984). Several factors have been found to bind these elements and other transcription binding sites *in vitro* and *in vivo* (figure 4). Transgenic studies have shown that the γ genes are autonomously silenced during development, even in the presence of the LCR (Dillon 1991). However, silencing can be abolished by specific mutations that have been found in individuals with HPFH (Berry 1992). Several of these mutations have been found in the CCAAT box region e.g. the -117 Greek HPFH (Collins 1985) and in an GC rich area at -200 bp of the transcription start site. Another element, called the stage selector element, is located 34-53 bp upstream of the transcription start site and has been reported to be involved in the γ to β switching process (SSE, Jane 1992). This sequence motif is a binding site for the stage selector protein (SSP) and has been implicated in the fetal activation of the γ genes. However, recent studies with transgenic mice having a deletion of the SSE reveal only slight differences in γ to β switching timing (Ristaldi in prep.).

The β globin promoter contains a TATA box, a CCAAT box and two CCAC boxes that are all required for proper expression. Studies on mouse and rabbit β globin gene promoters that have a promoter structure similar to the human β gene made clear that mutations or deletions in these elements severely lower the level of expression (Grosveld 1981,1982, Dierks 1981,1983). Transgenic mice with a β globin gene and promoter integrated in the mouse genome shows accurate developmental expression of the β globin gene albeit at very low levels (1%, Magram 1985, Townes 1985). In later experiments LCR gene constructs were used in order to get high levels of expression. Extensive studies with the TATA box region, using these type of constructs, showed that TFIID is the only factor binding this region *in vivo* despite nearby putative binding sites for Sp1 and GATA1 (Antoniou 1995). The CCAAT and CCAC boxes are essential for positioning of TFIID and both elements are required for a tissue specific high level of expression of the β globin gene (Antoniou 1990). Deletion of either the CCAAT or the CCAC boxes results in a 7-fold reduction of expression, whereas deletion of all CCAAT and CCAC boxes reduces expression 70 fold.

The significance of the proximal CACC box, located 90 bp upstream of the β globin transcription start site is most clearly shown by a naturally occurring mutation resulting in β thalassemia (Kulozik 1991). A red cell restricted transcription factor called erythroid krüppel like factor (EKLF, Miller 1993) specifically binds the proximal CCAC box of the β globin promoter *in vitro* (Feng 1994) and *in vivo* (chapter 3). EKLF knock out mice show that this protein is essential for expression of the murine and human adult β globin genes (Nuez 1995, Perkins 1995, chapter 3). Binding sites for EKLF are also found in HS3 of the murine and human LCR. Compound transgenic mice with a mutated EKLF consensus binding site in HS3 and concomitant mutations in the EKLF zinc finger that specifically bind mutated HS3 show that EKLF has *in vivo* opening properties on the LCR (Gillemans 1998). A recent report shows that EKLF co-operates with a chromatin remodelling complex E-RC1 to open up reconstituted chromatin *in vitro* at the β globin promoter creating a hypersensitive site (Armstrong 1998).

The organisation of proximal elements of the δ globin gene is similar to the β gene, nevertheless expression of this gene is only 2% compared the β gene. Recent reports show that this reduction in expression is the consequence of mutations in the CCAAT and CCAC boxes. Replacement of these mutated sequences with sequences of the β gene promoter reconstitutes expression levels comparable with the β gene expression (Donze 1996, Tang 1997).

The Locus Control Region

Initial evidence for the existence of additional elements necessary for proper expression of the β globin genes, besides promoter and gene sequences, came from a Dutch β thalassemia patient who has a 100 kb deletion starting 2.5 kb upstream of the β gene (van der Ploeg 1980, Kioussis 1983, Taramelli 1986). DNase I sensitivity studies showed that the mutant locus has an inactive chromatin configuration despite the presence of an intact β globin gene (Kioussis 1983). It was not clear however whether this inactive chromatin configuration was the result of the absence of *cis* regulatory elements or silencer elements juxtaposed to the β gene. Extensive studies with transgenic mice using β globin gene sequences alone indicated that *cis* regulatory elements were missing in these constructs (Magram 1985, Townes 1985, Kollias 1986). DNase I sensitivity studies upstream of the ε globin gene revealed five

erythroid specific developmentally stable hypersensitive sites (Tuan 1985, Grosveld 1987, Forrester 1987). Transgenic studies using 20 kb of sequence upstream of the ε gene, including all five hypersensitive sites, revealed tissue specific copy number dependent levels of expression of a linked β gene independent of the position of integration (Grosveld 1987). The area was called the locus control region (LCR) and many LCR's regulating other loci have been described since. A much smaller deletion spanning about 40 kb upstream of HS1 was reported in a Hispanic β thalassemic patient and provided more important data. Using somatic cell hybrids it was shown that like the Dutch thalassemia deletion the mutated locus is in an inactive chromatin configuration (Epner 1990). In contrast to the wildtype locus which replicates early in S phase in erythroid cells, the mutated locus replicates late in S phase. An origin of replication was identified <1 kb upstream of the β gene. It is still present in the Hispanic (and Dutch) locus suggesting a role for the LCR in mediating replication timing (Kitsberg 1993, Aladjem 1995). More recent studies suggest that the LCR is only indirectly involved in replication timing (Aladjem 1998). Early replication is probably the consequence of an open chromatin configuration which is not necessarily provided by the LCR. The deletion also showed that HS1 and another erythroid specific and developmentally stable hypersensitive site 3' of the locus (3'HS1) are not sufficient for proper expression of the globin genes *in vivo*.

A 6.5 kb micro-locus created by fusion of restriction fragments containing the core regions of hypersensitive sites 1-4 retains the functional activity of the LCR (Talbot 1989). The 200-300 bp core regions, contain several putative binding sites for erythroid specific and ubiquitously expressed proteins (figure 4.). All of the hypersensitive sites have GATA1, and NFE2 motifs as well as GT rich sequences that are putative binding sites for zinc finger proteins like Sp1 and EKLF. Most of these motifs are occupied by proteins *in vivo* in erythroid cells, as determined by DNase I footprinting analysis. It is nevertheless unclear if these *in vivo* footprints are the result of binding of the proteins, that have been identified by *in vitro* binding assays, for that specific site. Sp1 motifs for instance have been reported in several hypersensitive sites in the β globin LCR and promoters. However, *in vitro* differentiated Sp1 knockout cells still show transcription of embryonic β globin genes, suggesting that these Sp1 motifs are presumably occupied by different proteins (Marin 1997). Identical results were obtained with null mutants for both NFE2 subunits (Shivdasani 1995, Kotkow 1996), indicating that there is either a lot of redundancy or that the *in vivo* footprints are the result of factors that still have to be identified.

HS2 is the only hypersensitive site with enhancer activity in classical enhancer studies using transient transfection assays (Tuan 1989). This enhancer activity is dependent on NFE2 consensus sites (Ney 1990). In contrast stable transfection assays show that all HS1-4 have enhancer activity (Collis 1990). Linkage of a single HS1 and HS4 to a β globin gene resulted in a 10% expression level compared to the micro-locus, whereas expression levels up to 50% were obtained for HS2 and HS3. Identical results were reported using transgenic mice (Fraser 1990, 1993). All these studies made use of multi copy inserts of reporter constructs but subsequent studies with single copy inserts in transgenic mice showed that HS3 is the only hypersensitive site that has chromatin activation properties (Ellis 1996). HS3 is also the only hypersensitive site giving rise to γ and β expression in the fetal liver of transgenic mice when HS3 was linked two a γγδβ cosmid (Fraser 1993).

Analysis of transgenic mice with a full β globin locus with single deletions of hypersensitive site core regions showed that this results in a loss of copy number dependent and position independent expression of the transgene (Bungert 1995, Milot 1996). Loss of position independent expression is caused by a pericentromeric integration of the transgene that results in two different types of position effects (Milot 1996). Some lines have a heterocellular distribution of transgene expression

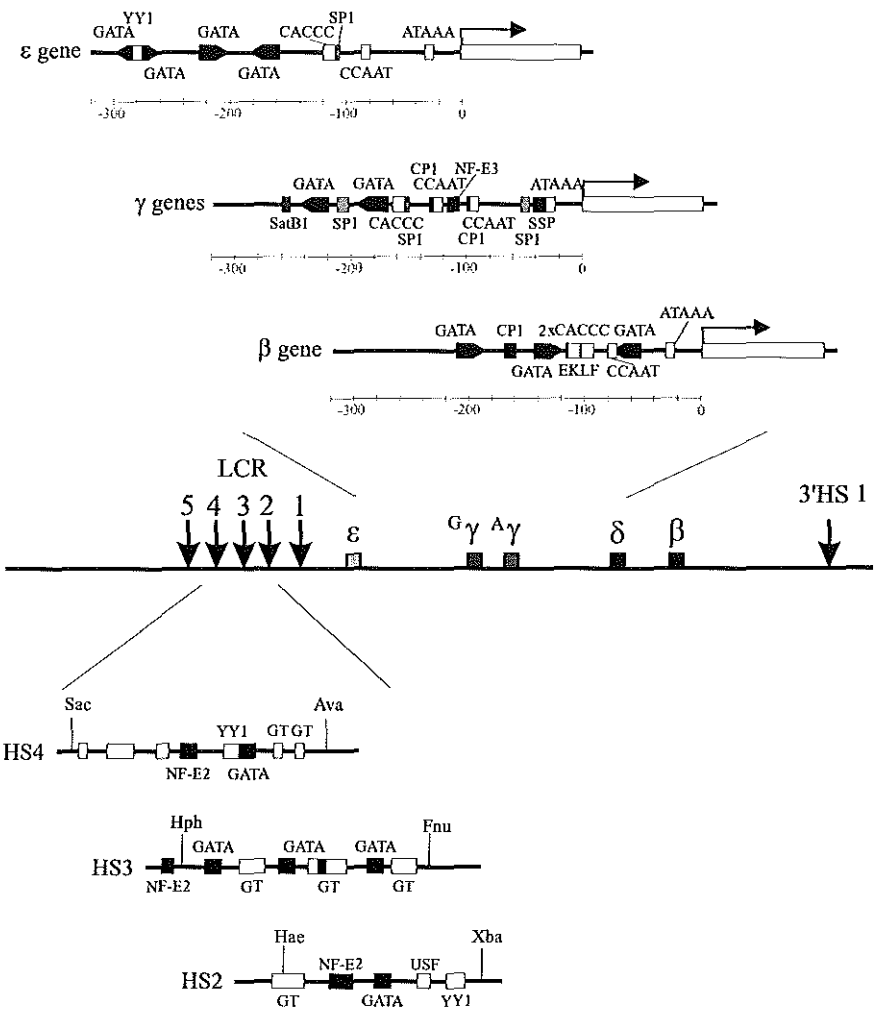


Figure 4. *Organization of putitive trans-acting factor binding sites in the LCR and promoter areas of the human β-like globin locus. Top panel shows the motifs reported in the promoters of the different β-like globin promoters. The bottom panel shows the motifs reported in hypersensitive sites 2,3 and 4 of the LCR.*

i.e. classical position effect variegation (PEV). Lines with severe PEV have a decreased sensitivity to DNase I which is probably the consequence of heterochromatinization of the transgene in part of the cells. Other lines displayed a pancellular expression pattern caused by a limited time span of expression of the transgene during the cell cycle. This new type of position effect was designated 'cell timing effect' (CTE), and seems to reflect changing chromatin states during the cell cycle (Milot pers. comm.). Loss of position independent expression upon deletion of hypersensitive sites of the LCR has been shown in different reports using transgenes of different sizes (Petersen 1996, Bungert 1999). Interestingly replacement of HS4 for HS3 shows that HS3 can fully complement HS4, whereas the converse replacement of HS3 for HS4 results in a loss of function of the LCR (Bungert 1995). All these results make clear that HS1-4 are all required for proper LCR function although some hypersensitive sites can be replaced by others. Non of the different hypersensitive sites have developmental or gene restricted specificity indicating that the individual hypersensitive sites act in concert as a holo-complex in order to activate distal genes (Milot 1996). Reik *et al.* (1998) recently showed that deletion of the LCR, in the context of the whole locus results in the complete abrogation of transcription but suprisingly showed that the locus with a deleted LCR remains in an active chromatin configuration. The results seem to contradict existing data, since loci with a deleted LCR found in the Hispanic and Dutch thalassemia patients are in an inactive chromatin configuration. Extensive studies with transgenic mice with small deletions of a single hypersensitive site of the LCR, in several cases leaving the regions intact that have been deleted in the Hispanic and Dutch thalassemia patients, support that the LCR is absolutely required for initiation of an active chromatin configuration. An explanation for this contradiction might be the fact that LCR deletion published by *Reik et al.* (1998) is performed in hybrid cells after transfer of human chromosome 11 into a murine erythroid cell line. This raises the question if a deletion prior to transfer to a erythroid cell line will give identical results, since epigenetic marks could already be established in case of an LCR deletion made in an erythroid cell line. The data however clearly shows that the LCR is absolutely required for high level transcription of the human globin genes.

HS5 is not required for proper LCR function but has been implicated to have a border or insulator function as has been shown for HS4 of the LCR of the chicken β globin locus (Chung 1993). The chicken β globin HS4 which shows similarity to human HS5, has insulator activity reported in erythroid cells and is also capable of protecting transgenes from PEV when assayed in *Drosophila* (Chung 1993, Li 1994). HS5 was originally described as a non erythroid specific hypersensitive site but later studies using transgenic mice made clear that HS5 is erythroid specific (Zafarana 1995). Transgenic studies show that HS5 can not block an upstream β globin gene from being activated by the LCR. Interestingly an additional construct using 15 kb of sequence upstream of HS5 does retain blocking activity which is absent in the subsequent 15 kb. Position effect assays confirm that HS5 is not capable of insulating a reporter gene from position effects (Zafarana 1995).

Elements involved in globin regulation

In addition to the hypersensitive sites in the LCR and promoter sequences several other sequences have been implicated in the regulation of β globin gene expression. Initial indications for such elements were usually based on erythroid specific hypersensitive sites located within these elements, and for most elements

enhancer activity was tested in transient or stable transfection assays. Recently most of these putative elements have been tested in transgenic mice with deletions made in the context of the whole locus. These data provided essential information for the understanding of the function of these elements *in vivo*.

The ε silencer element is a 225 bp sequence 177 bp upstream of the ε transcription start site. Initial studies using a transient transfection assay showed that this area has silencer capacities (Cao 1989). This was confirmed in transgenic mice and it was shown that deletion of this sequence from a μLCRε construct results in ε expression in definitive erythroid cells (Raich 1992). Contradictionary results were obtained when the element was deleted from yeast artificial chromosomes (YAC) containing a copy of the whole human β globin locus (Liu 1997). These mice do not express ε globin at all and fail to express γ globin in the embryonic blood cells, showing that the element should be regarded ε and γ silencer. Expression of γ and β globin in fetal liver and definitive erythrocytes is normal. The result might be an indication of the involvement of this element in the initiation of an embryonic domain containing at least the ε and γ globin genes. Another explanation could be that this element is involved in recruitment or positioning of the LCR and is required for the expression of different genes. The deleted sequence contains two inverted GATA1 motifs as well as a YY1 binding site that might mediate ε expression. The dual role of this element could partially be explained by this YY1 binding site, since a recent report shows that YY1 is a homologue of the *Drosophila* Polycomb group protein (PcG) pleiohomeotic (Brown 1998). Pleiohomeotic anchors PcG protein complexes to a PRE and similarly YY1 mediated silencing might invoke PcG proteins as well. It would therefore be interesting to monitor YY1 expression and activity throughout development. Interestingly in *Drosophila* several PRE's seem to overlap with Trithorax responsive elements (TRE), that are binding sites for Trithorax group proteins (Trx, Orlando 1998). Trx proteins contribute to strong expression of homeotic genes and a similar mechanism might explain the dual role of this ε globin silencer/enhancer element.

Similar results were obtained for a region located 4 kb upstream of the δ globin gene. This region of 700 bp in size has been of interest since it marks the difference between natural deletions giving rise to either a HPFH or a β thalassemia phenotype. The difference between the two phenotypes suggests that this region is involved in silencing of the γ genes, although the Corfu β thalassemia (with a deletion of this specific area) does not show elevated levels of γ expression. Transgenic mice with this region deleted confirmed the Corfu thalassemia phenotype and did not show any effect on the γ genes (Calzolari 1999). Like the Corfu thalassemia phenotype, the deletion results in an activation failure of the adult β gene and could, like the ε silencer/enhancer, play a role in positioning of the LCR. The area could however also be important in opening of a chromatin domain that is essential for β globin expression as will be discussed in chapter 6. Unlike previous results with globin transgenes that show a clear correlation between mRNA levels and transcription spots obtained with *in situ* hybridization (Wijgerde 1995, 1996, Milot 1996, Dillon 1997) both lines have decreased mRNA output levels compared to in situ hybridization signals. Run on analysis confirmed that the polymerase density on the β globin gene is indeed reduced in these lines.

A long stretch of pyrimidine rich sequences is located a few kb downstream of this element (~1 kb upstream of the δ globin gene). Deletion of this pyrimidine rich sequence in transgenic mice from a μLCR-^{A}γ-δ-β construct results in a delayed

switch from γ to β globin expression (O'Neill 1999). A complex, restricted to definitive hematopoietic cells, was purified that specifically binds this pyr sequence. The complex contains mammalian homologues of the yeast SWI/SNF subunits BAF57, INII, BAF60a and BAF170 suggesting that this element is involved in disrupting chromatin structure (Armstrong 1998).

Both the $^{A}\gamma$ and the β globin genes have 3' elements that have position and orientation independent enhancer properties in transient transfection assays (Wright 1984, Charnay 1984, Bodine 1987). The $^{A}\gamma$ enhancer is located 410 bp downstream of the poly adenylation site and contains two erythroid specific hypersensitive sites that have putative binding sites for GATA1 and different ubiquitous expressed proteins. *In vitro* studies implicated a role for this area in binding the nuclear scaffold (Cunningham 1994), although MAR/SAR activity was not mapped in that area in an earlier report (Jarman 1988). Transgenic mice with constructs containing the enhancer still show autonomous silencing (Dillon 1991), and a recent publication shows that the enhancer is essential for position independent expression (Stamatoyannopoulos 1997). Within the context of the whole locus however this enhancer is dispensable (Liu 1997). YAC mutant mice do not show a phenotype indicating that there is either enough redundancy in the locus or that the enhancer might be unimportant. In contradiction to the $^{A}\gamma$ enhancer the 3'β enhancer is essential for fetal and adult expression of the β globin gene when expression is assayed in YAC mutant mice. Interestingly deletion of the element (250 bp), located 600 bp downstream of the β globin gene does not result in an increase but even results in a decrease in the level of γ expression in the fetal liver (Liu 1997).

The murine globin loci

The mouse β globin locus is located on chromosome 7 and has an organisation very similar to the human β globin locus with 4 functional genes that are arranged in the order of their respective developmental expression pattern (Jahn 1980, figure 5). The murine counterpart of the human β globin LCR consists of 6 erythroid specific developmentally stable hypersensitive sites located upstream of these genes (Moon 1990, Hug 1992, Jimenez 1992, Bender 1998). Sequence comparison and subsequent transient and stable transfection experiments confirmed that human HS2 and HS3 have true murine homologues. Murine HS2 but not HS3 has enhancer activity in transient transfection experiments, whereas analogous to human HS2 and HS3 both murine hypersensitive sites have enhancer activity in stable transfection assays (Moon 1990, Hug 1992). Mice have two embryonically expressed globins, εy and βH1, that are down regulated when the fetal liver becomes the major source of erythropoiesis (Figure 5, Whitelaw 1990). The β^{maj} and β^{min} globin genes are transcriptionally active in embryonic blood cells at low levels and expression sharply increases in the fetal liver. The majority of adult β globin, about 80%, is produced by the β^{maj} gene. Interestingly deletion of the β^{maj} promoter and gene sequences results in a 50% increase in expression of the remaining β^{min} gene, suggesting that the murine β globin genes compete for activation by the LCR similar to the human β globin genes as will be discussed below (Skow 983, Curcio 1986, chapter 5). The strong homology of the human and murine β globin loci makes the murine locus a good model system for testing deletions of individual hypersensitive sites in there natural environment. To date all 6 murine hypersensitive sites have been deleted individually and tested for globin expression. Deletions of HS1, HS4 and HS5 result in a 10% decrease of adult β globin expression whereas HS2 and HS3

deletions show a 30% reduction in expression of the adult genes. In contrast to the deletion of HS2 that does not affect expression of the embryonic genes, deletion of HS3 does reduce expression of the embryonic εy gene by 20% and βH1 by 10%. Recently a remarkable set of results has been published concerning a deletion spanning all hypersensitive sites of the LCR (Epner 1998). Like the deletion of the human LCR deletion of the LCR from murine ES cells results in a severe reduction of globin gene expression but the locus with the deleted LCR is DNase I sensitive, after chromosome transfer into K562 cells. Differentiation of ES cells into embryonic bodies gave similar results and even showed that general DNase I sensitivity can be initiated upon *in vitro* differentiation. The remaining important question however is if identical results will be obtained after germline transmission of the deletion. Epigenetic marks, like polycomb and trithorax like proteins, are partially reset during germline transmission and can even be maintained for several germline transmissions (Cavalli 1998). Globin specific epigenetic marks could already be present in ES cells before deletion of the LCR. Importantly the results do show that the murine LCR is absolutely required for a high level of expression of the globin genes.

Erythroid lineage-specific transcription factors

Sequence comparison throughout the β globin locus reveals several globin specific highly conserved motifs usually located in promoters and hypersensitive sites of the LCR. *In vivo* footprinting analysis of these sequences showed that several of these motifs like GATA, CACC and CCAAT boxes are occupied by proteins. Subsequent band shift assays with these motifs indicated binding activity of different factors that have been extensively studied to unravel their individual roles in globin gene regulation. In this section I will mainly focus on the most studied erythroid specific proteins GATA1, EKLF and NFE2.

GATA1 was the first erythroid specific factor to be identified (Martin 1989). This zinc finger protein recognises the consensus sequence [T/A] GATA [A/G] that is found throughout the β globin locus, in promoter sequences and LCR hypersensitive sites, as well as in promoters of other erythroid specific genes. GATA1 is a member of the GATA family consisting of five other members of which only GATA1 and GATA2 are largely erythroid specific. Expression of GATA1 is not restricted to the erythroid lineage only but is also reported in megakaryocytes, eosinophils and mast cells. Initial transient transfection experiments showed that GATA1 regulates expression of several erythroid specific genes like α and β globin (Evans 1988, Wall 1988), EKLF (Crossley 1994) and the GATA1 gene self (Tsai 1991). Gene ablation of GATA1 in mice results in a block in definitive erythropoiesis (Pevny 1995), with an arrest of differentiation at the pro-erythroblast stage (Simon 1992). GATA2 knock out mice also die from anemia due to a decreased amount of hematopoietic precursors, suggesting that GATA2 acts earlier during differentiation (Tsai 1994). GATA1 interacts with several proteins like Sp1, EKLF (Merika and Orkin 1995, Gregory 1996), and YY1 and interestingly GATA1 binding motifs are frequently found in close proximity with motifs for one of these mentioned proteins. Recently another protein, Friend of GATA1 (FOG), has been reported to associate with GATA1 (Tsang 1997). Like $GATA1^{-/-}$ mice $FOG^{-/-}$ mice have severe anemia due to a partial arrest of differentiation at the pro-erythroblast stage (Tsang 1998). Interestingly GATA1 activity fluctuates during cell cycle progression and activity peaks in early and mid S phase and is low in G1, late S and

G2/M phase (Cullen 1997). This change in activity might be regulated by acetylation of the protein (Boyes 1998), since acetylation of GATA1 correlates with increased DNA binding activity. Recent data even suggest that GATA1 mediates cell cycle progression by regulating cyclin A2 levels (Whyatt submitted).

Murine Erythroid Kruppel-like factor (EKLF) was cloned using a subtractive hybridization screening of RNA transcripts from erythroid specific MEL cells and a myelomonocytic cell line (Miller 1993). EKLF is an erythroid specific protein expressed at all developmentally erythropoietic stages and is besides the erythroid lineage expressed in mast cells albeit at lower levels. The human homologue of EKLF has been characterised and shows extensive sequence similarity and highly conserved structural properties with murine EKLF (Bieker 1996). EKLF contains a proline rich transcription activation domain and three Sp1 like zinc fingers that specifically bind the sequence CCACACCCT (Feng 1994), found in the murine and human adult β globin promoters and other sites in the locus like HS3 of the LCR. Initial evidence for the importance of this sequence for adult β globin expression, came from β thalassemia patients having mutations in this specific sequence of the β globin promoter (Kulozik 1991). Reporter constructs with these mutations were unable to be activated by EKLF (Feng 1994). Although EKLF is expressed during development in the yolk sac, co-transfection experiments with EKLF and either a γ or β globin reporter construct show that EKLF preferentially activates β globin gene expression (Donze 1995). This was confirmed by genetic ablation studies of EKLF in mice, using LacZ-neo or neo cassettes (Nuez 1995, Perkins 1995). EKLF$^{-/-}$ mice appear normal during embryonic development but die at day 12-14, as a consequence of severe anemia. Adult murine β^{maj} and β^{min} globin expression is severely reduced in these mice, whereas the absence of EKLF does not affect α globin and embryonic β globin expression. The results indicate the importance of EKLF in globin gene switching and additional data will be presented in chapter 3 concerning human β globin switching in compound transgenic mice having the complete human β globin locus crossed into a EKLF$^{+/-}$ or EKLF$^{-/-}$ background.

Recently it has been shown that EKLF requires a SWI/SNF related chromatin remodelling complex called, EKLF coactivator remodelling complex 1 (E-RC1, Armstrong 1998), in order to form a DNase I hypersensitive and transcriptionally active β globin promoter *in vitro*. E-RC1 was initially isolated from MEL cells but could also be purified from HELA cells. Hypersensitive site formation and transcription activation of a β globin gene from a chromatin template is ATP dependent and requires both EKLF and E-RC1. Deletion of the transactivation domain of EKLF results in abrogation of transcription but hypersensitive site formation remains unaffected, showing that hypersensitive site formation itself is not sufficient for, or a consequence of, transcription of a globin gene. Interestingly another recent report shows that EKLF can be phosphorylated *in vitro* by casein kinase II (Ouyang 1998). Mutation analysis of the phosphorylation site results in a downregulation of EKLF transactivation *in vivo*. Like GATA1, EKLF has been found to be acetylated *in vivo* (Zhang 1998) and *in vitro*. This study indicates that this posttranslational modification of EKLF is mediated by CBP and P300. Both modifications have been implicated in the downregulation of EKLF during early development, since EKLF is present in the yolk sac without the presence of a hypersensitive site in the adult β globin promoter at that stage. However constructs with the β globin gene placed in close proximity to the LCR or in the position of the ε globin gene in whole locus constructs express high levels of β globin in embryonic

blood (Dillon 1997), suggesting the possibility that EKLF is already active at that stage. Additional evidence for embryonic activity of EKLF came from compound transgenic mice with constructs consisting of HS3 or the HS3 core region linked to a heterologous promoter that were bred into an EKLF$^{-/-}$ background (Tewari 1998). Abrogation of EKLF leads to a severe downregulation of expression of the construct in the embryonic blood and fetal liver and shows that EKLF acts on HS3 and is already active in the embryo. EKLF binds the HS3 core region as was shown by compound transgenic mice with mutations in the EKLF binding site and concomitant mutations in the EKLF zincfinger that enables mutant EKLF to bind the mutated binding site (Gillemans 1998). Mutated EKLF specifically acts on the mutated EKLF motif whereas mutations in Sp1, another candidate for binding this sequence, did not result in activation of the downstream β promoter.

NFE2is a leucine zipper protein that specifically binds AP1 motifs, [T/C] GCTGA [C/G] TCA [T/C], that can be found in all hypersensitive sites of the human β globin LCR and the αMRE. The protein is a heterodimer of an erythroid restricted 45kD protein p45 and the ubiquitous expressed 18 kD protein p18. The p45 subunit is a member of a subfamily of bZIP proteins like Nrf1/LCRF1 and Nrf2, whereas p18 that lacks a transactivation domain, belongs to the Maf family (for review see Baron 1996). NFE2 motifs found in HS2 of the human β globin LCR have been shown to be essential for HS2 function (Ney 1990, Talbot 1990). *In vitro* studies with chromatin templates show that NFE2 is capable of disrupting nucleosomes and that nucleosomal disruption by NFE2 is a prerequisite for GATA1 to bind nearby sites (Armstrong 1996). NFE2 has also been implicated in LCR mediated long range activation of the β globin gene via an interaction of NFE2 and TAF$_{II}$130 (Amrolia 1997). Although NFE2 is required for high levels of β globin expression in MEL cells, p45 and p18 null mutations in mice do not have a severe effect on β globin expression (Shivdasani 1995, Kotkow 1996), suggesting a high level of redundancy with other proteins taking over NFE2 function or that the NFE2 motifs could be dispensable within the context of the whole locus. There is no indication yet that supports redundant activity of different proteins for p45 action since null mutant mice for p45 do not show new complexes binding the NFE2 motif in a gel shift assay (Shivdasani 1995). Knock out studies with the family members of p45, Nrf1 and Nrf2 and even compound p45-Nrf2 null mutants do not show severe effects on erythropoiesis (Chan 1998, Chan 1996, Martin 1998), although it remains to be seen what happens when both p45 and Nrf1 genes are knocked out. In contradiction to the p45 null mutant p18 knockout mice still show NFE2 gel shift activity suggesting that this protein is substituted by another Maf family member (Kotkow 1996).

Globin gene switching in transgenic mice

Transgenic mice have been important for the understanding of switching of the β globin genes during development. The first experiments using transgenic mice were published in the early 80's, with the rabbit and human β globin gene sequences among the first to be injected in mouse oocytes (Constantini 1981, Steward 1982). These initial experiments made clear that germline transmission of foreign DNA in the mouse genome is possible, however globin expression was absent or expression was reported in inappropriate tissues (Lacy 1983). Removal of vector sequences finally resulted in tissue specific developmentally specific expression of the adult β gene (Magram 1985, Townes 1985). Identical results were obtained for the γ globin genes albeit all authors reported very low levels of expression of the transgene

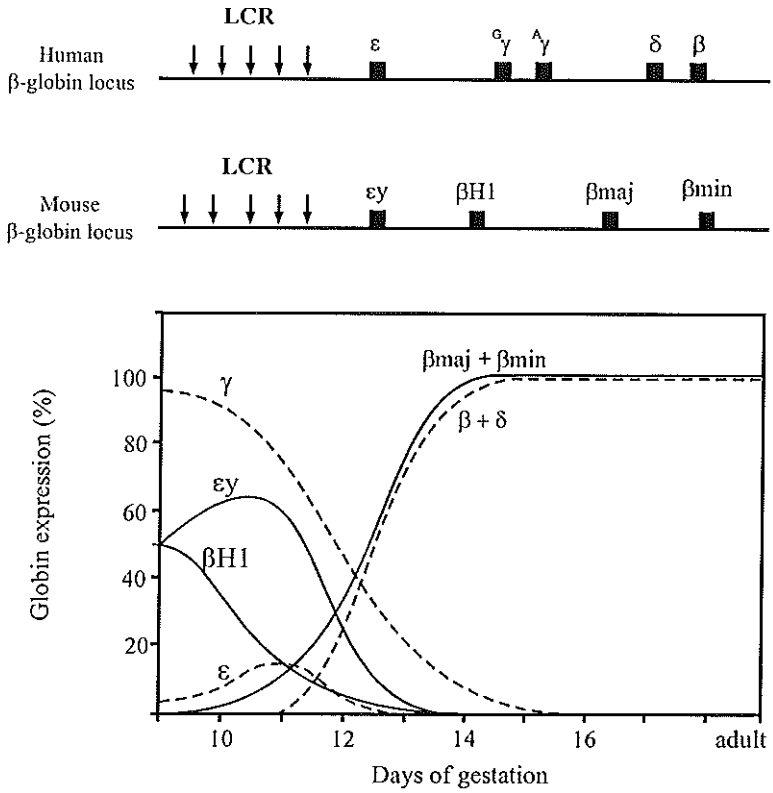


Figure 5. *Organization and expression of the different β-like globin genes of the human and murine β-like globin loci.*

(Chada 1986, Kollias 1986). Studies with the ε globin gene did not reveal any ε expression during development (Shih 1990). In subsequent studies LCR or μLCR-gene constructs were introduced in the mouse germline. Addition of the LCR results in position independent and copy number dependent tissue specific expression of the linked gene (Grosveld 1987). Both the ε and the γ globin genes are autonomously regulated during development in micro and full LCR constructs (Raich 1990, Shih 1990, Fraser 1993 and Dillon 1991). The β globin gene however is precociously expressed in embryonic blood and remains expressed at high levels in the fetal liver and adult (Enver 1990). The results suggest that the genic sequences and sequences nearby genes contain most of the information necessary for proper expression during development, with stage specific silencing being the important parameter for the regulation of individual genes. Attempts to monitor globin gene switching using constructs without the LCR resulted in very low expression levels like transgenic mice with single globin sequences (Fraser 1993, Starck 1994), although Starck *et al.* (1994) reported correct developmental switching. The first full locus transgenics were constructed by ligation of two cosmids, together comprising 70 kb of human β globin genomic sequence (Strouboulis 1992a, 1992b). These mice display high levels

of expression, comparable to expression levels of the murine globin genes, and developmental switching, showing that transgenic mice represent a good model system for studying switching of the human β globin genes. The ε and γ globin genes are expressed in embryonic blood, both the γ and β globin genes in the fetal liver and the β globin gene alone is expressed in the adult (figure 5). In contradiction to the expression pattern in human, transgenic ε globin expression increases during development in embryonic blood whereas γ expression is already high early in embryonic blood and declines during development. Another difference is found in the switch from γ to β globin expression. In human this switch takes place around birth whereas transgenic mice have a relative accelerated switching pattern taking place early during erythropoiesis in the fetal liver. Interestingly near accurate developmental expression of the fetal γ gene is still observed in these mice despite the fact that there are no equivalent murine fetal globin genes. Identical results have been reported for YAC transgenic mice with 150 kb up to 248 kb of genomic sequence (Gaensler 1993, Peterson 1993). Immunofluorescence studies showed that single fetal liver cells of transgenic mice stain positive for both γ and β globin proteins (Fraser 1993). This was confirmed by single cell PCR and indicated that during switching different globin genes can be activated within the same cell (Fukuwara 1994, 1995). In successive RNA *in situ* hybridization studies developmental expression of the β globin genes in single copy full locus transgenic mice (Strouboulis 1992a) was examined using intron specific oligo's (Wijgerde 1995). This powerful technique allows the detection of actively transcribing or recently active genes since the half-life of a β globin primary transcript is shorter than 4-5 minutes. Abrogation of transcription using actinomycin-D results in complete disappearance of transcription foci in 7 minutes, indicating that transcription foci represent genes that were transcribing within the timespan of 7 minutes (chapter 4). Hybridization of 12.5 day fetal liver cells of mice heterozygous for the human β globin locus using γ and β globin specific oligo's revealed mostly cells with single γ or β transcription foci (Wijgerde 1995). A minority of cells had foci (15%) with both γ and β globin signals on the same allele. Homozygous 12.5 day fetal liver cells displayed all possible combinations of foci, showing that the different loci are not regulated synchronously within a cell. Interestingly detection for γ and β transcription foci in conjunction with detection for β mRNA in the cytoplasm of heterozygous 12.5 day fetal liver cells resulted in a subset of cells that have a β globin cytoplasmic halo with a single γ globin transcription spot. It was therefore concluded that the switching process is not a progressive switch from transcription of γ globin genes followed by transcription of the β globin gene, but that it rather is a dynamic process with different globin genes alternatively transcribed during development (Wijgerde 1995).

Competition of the globin genes

Transgenic mice with constructs harbouring a single β globin gene linked to the LCR express β globin in the yolk sac (Enver 1990). Addition of a γ globin gene in between the LCR and the β gene however abolishes this embryonic β globin expression (Enver 1990, Hanscombe 1991), restoring correct developmental expression. Similar results were obtained using the chicken β globin locus as a model system (Choi 1988). From these observations and several other studies with different LCR gene constructs it became apparent that the different globin genes compete for the LCR. Expression of the different globin genes is dependent on the relative

position of the gene to the LCR (Hanscombe 1991). The competition model implies that different genes compete for an interaction with the LCR and suggests that genes positioned closer to the LCR have a higher change of interacting with the LCR. Stage-specific elements determine which genes can be active at a certain stage during development. These initial studies described results obtained with multi copy transgenics with relatively small constructs, making it difficult to extract how relative positioning of the different genes to the LCR is related to gene expression. In successive experiments the importance of gene position in the context of the whole locus was examined by placing a marked β globin gene at different positions in the locus (Dillon 1997). The advantage of this strategy is that the two β genes being examined have identical promoter strengths. Two different artificial loci were generated with the marked β globin gene ($^{M}\beta$) replacing either the ε gene or inserted 5' of the δ globin gene. Lines with the $^{M}\beta$ gene in the ε position show high levels of $^{M}\beta$ expression throughout development. In this position the $^{M}\beta$ gene is active in the embryo and suppresses the wildtype β gene in adult erythrocytes. Mouse lines with the $^{M}\beta$ gene in the distal position do not have embryonic $^{M}\beta$ expression and $^{M}\beta$ expression in the adult is 2 times higher than the partially suppressed wild type β gene. Thus changes in the relative distance to the LCR result in a change in the frequency of LCR gene interactions. These results are consistent with earlier reports and clearly show that the relative distance to the LCR is a key parameter for globin gene expression. Subsequent data concerning the influence of trans-acting factors like EKLF on competition will be discussed in chapter 3.

Chromatin domains and epigenetics

Evidence is accumulating that suggest that the activation of globin genes and switching is controlled through more elaborate means than the availability of developmentally specific trans-acting factors and gene competition for the LCR. It has long been proposed that the globin locus may be divided into distinct subdomains based primarily on the analysis of naturally occurring mutations (Huisman 1974, Bernards 1980, Collins and Weisman 1984) and that switching is controlled in part by 'opening and closing' these domains. Expression analysis of somatic cell hybrids created by fusing γ-expressing human fetal erythroid cells with mouse erythroleukemia (MEL) cells indicates that the human globin locus remembers the *in vivo* expression pattern (Papayannopoulou 1986). Remaining γ globin gene expression is maintained through several rounds of replication in the hybrid adult MEL cell environment. This observation becomes significant in light of the fact that the transfection of naked DNA constructs into MEL cells does not result in developmentally regulated expression and fusion of fetal nonerythroid cells with MEL cells results in only adult β globin expression. These results have suggested that epigenetic modification of the globin locus is acquired during development in erythroid cells and is stable through several hybrid cell divisions before eventually switching to β globin expression after 10 weeks in culture (Papayannopoulou 1986, Stanworth 1995). Another rather surprising set of results may provide further evidence for domain organisation of the β globin locus. Liu *et al.* (1997) analysed expression in human β-globin yeast artificial chromosome-containing transgenic mice with a small deletion of the ε silencer. Contrary to expectations, deletion of the ε silencer did not result in persistent expression of the ε globin gene as has been shown in partial LCR constructs, which contained only a single modified ε globin gene, but instead resulted in failure to activate the ε globin gene. Interestingly

expression of γ globin was also severely reduced in embryonic blood but was normal in fetal liver cells. Similar results were obtained by Calzolari *et al.* (1999). Deletion of an area that has been implicated in developmental silencing of the γ globin genes based on patient material, resulted in normal silencing of the γ globin genes but showed an activation failure of the adult β globin gene. These results could be interpreted to indicate that these regions are important for activation of overlapping developmentally specific domains, which include the ε and γ globin genes in the embryo and the δ and β globin genes in the fetal liver and adult. Recently data have been published concerning the deletion of an LCR in an erythroid background (Reik 1998) which show that information for the propagation of active chromatin is retained within the locus. The elements mentioned above could play a role in this. However, the Dutch and Hispanic thalassemia patient data and data concerning transgenic mice with small deletions of single hypersensitive sites in the LCR clearly show that the LCR is absolutely required for the initial activation of the locus. Work will be presented in chapter 6 providing further evidence that elements outside the LCR are required for the opening of active chromatin domains and are therefore important for position independent activity in the context of the full locus.

Models for globin gene activation

Three models have been proposed for the mechanism of LCR activation and developmental regulation of the β globin genes (figure 6). They can be described as an accessibility model (Martin 1996), a tracking model (Herendeen 1992, Tuan 1992) and a looping model (Ptashne 1988, Mueller-Storm 1989, Bickel and Pirotta 1990). All three models accept that the LCR is involved in the initial chromatin activation of the locus.

The accessibility model proposes that the singular role of the LCR is to open the chromatin structure of the globin domain, which allows stage-specific factors to gain access to gene-local regulatory elements. The globin genes would then be activated in a stochastic manner, switching on and off in response to the changing trans-acting factor environment leading to a progressive increase in adult β globin expression during development. This model explains the importance of gene order in

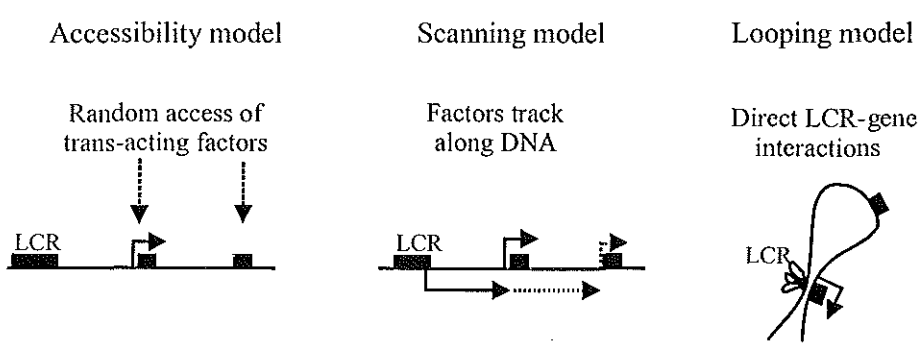


Figure 6. *The different models that have been proposed for explaining activation of the different β globin genes by the LCR.*

competition and single-gene activation by proposing that transcriptional interference, caused by transcriptional activity of an upstream gene in the locus, could decrease the probability of transcription of a downstream gene. However transcriptional interference does not explain the result obtained with mouse lines having a second marked β globin gene inserted at different positions in the locus (Dillon 1997). Transcriptional interference predicts a more severe interference when the marked β gene is moved closer to the downstream wildtype β gene, however Dillon *et al.* (1997) showed that moving the marked β gene closer to the wildtype β gene results in an increase of expression of the downstream wild type gene.

The tracking model proposes that the LCR and, in particular, HS2 is an entry point for polymerase. Polymerases then track along the DNA in a unidirectional manner toward a promoter that is rendered activatable by the trans-acting environment (Herendeen 1992, Tuan 1992). This model could account for the competitive effect of gene order with respect to the LCR and single-gene transcription. However, it cannot adequately explain the observation of alternating transcription without further conditions (Wijgerde 1995). Nor can it explain differences obtained by varying distances between genes without changing gene order (Dillon 1997).

The looping model proposes that the LCR forms dynamic, direct chromatin interactions with gene-local regulatory elements of a single gene via looping to initiate transcription. This model accounts for alternating transcription of single genes and differences in the competitive ability of genes in the locus governed in part by gene order and the changing trans-acting factor environment. The model is supported by *in situ* hybridization data that show alternate transcription of γ and β globin expression and explains why a gene moved closer toward the LCR has an increased competitive advantage over a distal gene. The crucial difference between the looping model and both the accessibility and scanning model is that the looping model is the only model that allows single gene activation whereas the other models allow co-activation of multiple genes *in cis*. This question will be addressed in chapters 4 and 5.

References

Aladjem M.I., Groudine M., Brody L.L., Dieken E.S., Fournier R.E., Wahl G.M., Epner E.M. 1995. Participation of the human beta-globin locus control region in initiation of DNA replication. *Science* **270**:815-9.

Aladjem M.I., Rodewald L.W., Kolman J.L., Wahl G.M. 1998. Genetic dissection of a mammalian replicator in the human beta-globin locus. *Science* **281**:1005-9.

Allan M., Lanyon W.G., Paul J. 1983. Multiple origins of transcription in the 4.5 Kb upstream of the epsilon-globin gene. *Cell* **35**:187-97.

Amrolia PJ, Ramamurthy L, Saluja D, Tanese N, Jane SM, Cunningham JM 1997. The activation domain of the enhancer binding protein p45NF-E2 interacts with TAFII130 and mediates long-range activation of the alpha- and beta-globin gene loci in an erythroid cell line. *Proc Natl Acad Sci U S A* **94**:10051-6.

Anagnou N.P., Papayannopoulou T., Stamatoyannopoulos G., Nienhuis A.W. 1985. Structurally diverse molecular deletions in the beta-globin gene cluster exhibit an identical phenotype on interaction with the beta S-gene. *Blood* **65**:1245-51.

Antoniou M., Grosveld F. 1990. Beta-globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev* **4**:1007-13.

Antoniou M., de Boer E., Spanopoulou E., Imam A., Grosveld F. 1995. TBP binding and the rate of transcription initiation from the human beta-globin gene. *Nucleic Acids Res* **23**:3473-80.

Custodio N., Carmo-Fonseca M., Geraghty F., Pereira H.S., Grosveld F., Antoniou M. 1999. Inefficient processing impairs release of RNA from the site of transcription. *EMBO J* **18**:2855-2866.

Armstrong J.A., Bieker J.J., Emerson B.M. 1998. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* **95**:93-104.

Ashe H.L., Monks J., Wijgerde M., Fraser P., Proudfoot N.J. 1997. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11**:2494-509.

Baron M.H. 1996. Developmental regulation of the vertebrate globin multigene family. *Gene Expr* **6**:129-37.

Bender M.A., Reik A., Close J., Telling A., Epner E., Fiering S., Hardison R., Groudine M. 1998. Description and targeted deletion of 5' hypersensitive site 5 and 6 of the mouse beta-globin locus control region. *Blood* **92**:4394-403.

Bernards R., Flavell R.A. 1980. Physical mapping of the globin gene deletion in hereditary persistence of foetal haemoglobin (HPFH). *Nucleic Acids Res* **8**:1521-34.

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

Bickel S., Pirrotta V. 1990. Self-association of the *Drosophila* zeste protein is responsible for transvection effects. *EMBO J* **9**:2959-67.

Bieker J.J. 1996. Isolation, genomic structure, and expression of human erythroid Kruppel-like factor (EKLF). *DNA Cell Biol* **15**:347-52.

Blom van Assendelft G., Hanscombe O., Grosveld F., Greaves D.R. 1989. The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell* **56**:969-77.

Bodine D.M., Ley T.J. 1987. An enhancer element lies 3' to the human A gamma globin gene. *EMBO J* **6**:2997-3004.

Boyes J., Byfield P., Nakatani Y., Ogryzko V. 1998. Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* **396**:594-8.

Bresnick E.H., Felsenfeld G. 1994. Dual promoter activation by the human beta-globin

locus control region. *Proc Natl Acad Sci U S A* 91:1314-7.
Brown J.L., Mucci D., Whiteley M., Dirksen M.L., Kassis J.A. 1998. The *Drosophila* Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell* 1:1057-64.
Bungert J., Dave U., Lim K.C., Lieuw K.H., Shavit J.A., Liu Q., Engel J.D. 1995. Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev* 9:3083-96.
Bungert J., Tanimoto K., Patel S., Liu Q., Fear M., Engel J.D. 1999. Hypersensitive site 2 specifies a unique function within the human beta-globin locus control region to stimulate globin gene transcription. *Mol Cell Biol* 19:3062-72.
Calzolari R., McMorrow T., Yannoutsos N., Langeveld A., Grosveld F. 1999. Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and deltabeta-thalassemia affects beta- but not gamma-globin gene expression. *EMBO J* **18**:949-58.
Camaschella C., Serra A., Gottardi E., Alfarano A., Revello D., Mazza U., Saglio G. 1990. A new hereditary persistence of fetal hemoglobin deletion has the breakpoint within the 3' beta-globin gene enhancer. *Blood* **75**:1000-5.
Cao S.X., Gutman P.D., Dave H.P., Schechter A.N. 1989. Identification of a transcriptional silencer in the 5'-flanking region of the human epsilon-globin gene. *Proc Natl Acad Sci U S A* **86**:5306-9.
Cavalli G., Paro R. 1998. The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**:505-18.
Chada K., Magram J., Costantini F. 1986. An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature* **319**:685-9.
Chan K., Lu R., Chang J.C., Kan Y.W. 1996. NRF2, a member of the NFE2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development. *Proc Natl Acad Sci U S A* **93**:13943-8.
Chan J.Y., Kwong M., Lu R., Chang J., Wang B., Yen T.S., Kan Y.W. 1998. Targeted disruption of the ubiquitous CNC-bZIP transcription factor, Nrf-1, results in anemia and embryonic lethality in mice. *EMBO J* **17**:1779-87.
Charnay P., Treisman R., Mellon P., Chao M., Axel R., Maniatis T. 1984. Differences in human alpha- and beta-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. *Cell* **38**:251-63.
Cho H., Orphanides G., Sun X., Yang X.J., Ogryzko V., Lees E., Nakatani Y., Reinberg D. 1998. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol* **18**:5355-63.
Choi O.R., Engel J.D. 1988. Developmental regulation of beta-globin gene switching. *Cell* **55**:17-26.
Chung J.H., Whiteley M., Felsenfeld G. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **74**:505-14.
Collins F.S., Weissman S.M. 1984. The molecular genetics of human hemoglobin. *Prog Nucleic Acid Res Mol Biol* **31**:315-462.
Collins F.S., Metherall J.E., Yamakawa M., Pan J., Weissman S.M., Forget B.G. 1985. A point mutation in the A gamma-globin gene promoter in Greek hereditary persistence of fetal haemoglobin. *Nature* **313**:325-6.
Collis P., Antoniou M., 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* 9:233-40.

Costantini F., Lacy E. 1981. Introduction of a rabbit beta-globin gene into the mouse germ line. *Nature* **294**:92-4.
Crossley M., Tsang A.P., Bieker J.J., Orkin S.H. 1994. Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. *J Biol Chem* **269**:15440-4.
Cullen M.E., Patient R.K. 1997. GATA-1 DNA binding activity is down-regulated in late S phase in erythroid cells. *J Biol Chem* **272**:2464-9.
Cunningham J.M., Purucker M.E., Jane S.M., Safer B., Vanin E.F., Ney P.A., Lowrey C.H., Nienhuis A.W. 1998. The regulatory element 3' to the A gamma-globin gene binds to the nuclear matrix and interacts with special A-T-rich binding protein 1 (SATB1), an SAR/MAR-associating region DNA binding protein. *Blood* **84**:1298-308.
Curcio M.J., Kantoff P., Schafer M.P., Anderson W.F., Safer B. 1986. Compensatory increase in levels of beta minor globin in murine beta-thalassemia is under translational control. *J Biol Chem* **261**:16126-32.
Dierks P., van Ooyen A., Mantei N., Weissmann C. 1981. DNA sequences preceding the rabbit beta-globin gene are required for formation in mouse L cells of beta-globin RNA with the correct 5' terminus. *Proc Natl Acad Sci U S A* **78**:1411-5.
Dierks P., van Ooyen A., Cochran M.D., Dobkin C., Reiser J., 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., Grosveld F. 1991. Human gamma-globin genes silenced independently of other genes in the beta-globin locus. *Nature* **350**:252-4.
Dillon N., Trimborn T., Strouboulis J., Fraser P., Grosveld F. 1997. The effect of distance on long-range chromatin interactions. *Mol Cell* 1:131-9.
Donze D., Townes T.M., Bieker J.J. 1995. Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching. *J Biol Chem* **270**:1955-9.
Donze D., Jeancake P.H., Townes T.M. 1996. Activation of delta-globin gene expression by erythroid Krupple-like factor: a potential approach for gene therapy of sickle cell disease. *Blood* **88**:4051-7.
Ellis J., Tan-Un K.C., Harper A., Michalovich D., Yannoutsos N., Philipsen S., Grosveld F. 1996. A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *EMBO J* **15**:562-8.
Enver T., Raich N., Ebens A.J., Papayannopoulou T., Costantini F., Stamatoyannopoulos G. 1990. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* **344**:309-13.
Epner E., Reik A., Cimbora D., Telling A., Bender M.A., Fiering S., Enver T., Martin D.I., Kennedy M., Keller G., Groudine M. 1998. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell* **2**:447-55.
Evans T., Reitman M., Felsenfeld G. 1988. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc Natl Acad Sci U S A* **85**:5976-80.
Feng W.C., Southwood C.M., Bieker J.J. 1994. Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* **269**:1493-500.
Forrester W.C. Takegawa S., Papayannopoulou T., Stamatoyannopoulos G., Groudine M. 1987. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res* **15**:10159-77.
Forrester W.C., Epner E., Driscoll M.C., Enver T., Brice M., Papayannopoulou T.,

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* 4:1637-49.
Fraser P., Hurst J., Collis P., Grosveld F. 1990. DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res* 18:3503-8.
Fraser P., Pruzina S., Antoniou M., Grosveld F. 1993. Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev* 7:106-13.
Furukawa T., Zitnik G., Leppig K., Papayannopoulou T., Stamatoyannopoulos G. 1994. Coexpression 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* 83:1412-9.
Furukawa T., Navas P.A., Josephson B.M., Peterson K.R., Papayannopoulou T., Stamatoyannopoulos G. 1995. Coexpression of epsilon, G gamma and A gamma globin mRNA in embryonic red blood cells from a single copy beta-YAC transgenic mouse. *Blood Cells Mol Dis* 21:168-78.
Gaensler K.M., Kitamura M., 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 U S A* 90:11381-5.
Gibbons R.J., Picketts D.J., Villard L., Higgs D.R. 1995. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* 80:837-45.
Gillemans N., Tewari R., Lindeboom F., Rottier R., de Wit T., Wijgerde M., Grosveld F., Philipsen S. 1998. Altered DNA-binding specificity mutants of EKLF and Sp1 show that EKLF is an activator of the beta-globin locus control region in vivo. *Genes Dev* 12:2863-73.
Gregory R.C., Taxman D.J., Seshasayee D., Kensinger M.H., Bieker J.J., Wojchowski D.M. 1996. Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. *Blood* 87:1793-801.
Grosveld F., van Assendelft G.B., Greaves D.R., Kollias G. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51:975-85.
Grosveld F. 1999. Activation by locus control regions? *Curr Opin Genet Dev* 9:152-157.
Grosveld G.C., Shewmaker C.K., Jat P., Flavell R.A. 1981. Localization of DNA sequences necessary for transcription of the rabbit beta-globin gene in vitro. *Cell* 25:215-26.
Grosveld G.C., de Boer E., Shewmaker C.K., Flavell R.A. 1982. DNA sequences necessary for transcription of the rabbit beta-globin gene in vivo. *Nature* 295:120-6.
Groudine M., Kohwi-Shigematsu T., Gelinas R., Stamatoyannopoulos G., Papayannopoulou T. 1983. Human fetal to adult hemoglobin switching: changes in chromatin structure of the beta-globin gene locus. *Proc Natl Acad Sci U S A* 80:7551-5.
Hanscombe O., Whyatt D., Fraser P., Yannoutsos N., Greaves D., Dillon N., Grosveld F. 1991. Importance of globin gene order for correct developmental expression. *Genes Dev* 5:1387-94.
Hardison R.C. 1996. A brief history of hemoglobins: plant, animal, protist, and bacteria. *Proc Natl Acad Sci U S A* 93:5675-9.
Hebbes T.R., Clayton A.L., Thorne A.W., Crane-Robinson C. 1994. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J* 13:1823-30.
Herendeen D.R., Kassavetis G.A., Geiduschek E.P. 1992. A transcriptional enhancer

whose function imposes a requirement that proteins track along DNA. *Science* **256**:1298-303.
Higgs D.R. 1993. Alpha-Thalassaemia. *Baillieres Clin Haematol* **6**:117-50.
Hug B.A., Moon A.M., Ley T.J. 1992. Structure and function of the murine beta-globin locus control region 5' HS-3. *Nucleic Acids Res* **20**:5771-8.
Huisman T.H., Schroeder W.A., Efremov G.D., Duma H., Mladenovski B., Hyman C.B., Rachmilewitz E.A., Bouver N., Miller A., Brodie A., Shelton J.R., Shelton J.B., Apell G. 1974. The present status of the heterogeneity of fetal hemoglobin in beta-thalassemia: an attempt to unify some observations in thalassemia and related conditions. *Ann N Y Acad Sci* **232**:107-24.
Jahn C.L., Hutchison C.A., Phillips S.J., Weaver S., Haigwood N.L., Voliva C.F., Edgell M.H. 1980. DNA sequence organization of the beta-globin complex in the BALB/c mouse. *Cell* **21**:159-68.
Jane S.M., Ney P.A., Vanin E.F., Gumucio D.L., Nienhuis A.W. 1992. Identification of a stage selector element in the human gamma-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the beta-promoter. *EMBO J* 11:2961-9.
Jarman A.P., Higgs D.R. 1988. Nuclear scaffold attachment sites in the human globin gene complexes. *EMBO J* 7:3337-44.
Jimenez G., Gale K.B., Enver T. 1992. The mouse beta-globin locus control region: hypersensitive sites 3 and 4. *Nucleic Acids Res* **20**:5797-803.
Kadosh D., Struhl K. 1998. Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo. *Mol Cell Biol* **18**:5121-7.
Kioussis D., Vanin E., deLange T., Flavell R.A., Grosveld F.G. 1983. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* **306**:662-6.
Kitsberg D., Selig S., Keshet I., Cedar H. 1993. Replication structure of the human beta-globin gene domain. *Nature* **366**:588-90.
Kotkow K.J., Orkin S.H. 1996. Complexity of the erythroid transcription factor NF-E2 as revealed by gene targeting of the mouse p18 NF-E2 locus. *Proc Natl Acad Sci U S A* 93:3514-8.
Kulozik A.E., Yarwood N., Jones R.W. 1988. The Corfu delta beta zero thalassemia: a small deletion acts at a distance to slectively abolish beta globin gene expression. *Blood* 71:457-6.
Kulozik A.E., Bellan-Koch A., Bail S., Kohne E., Kleihauer E. 1991. Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. *Blood* 77:2054-8.
Kollias G., Wrighton N., Hurst J., 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.
Lacy E., Roberts S., Evans E.P., Burtenshaw M.D., Costantini F.D. 1983. A foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and expression in inappropriate tissues. *Cell* 34:343-58.
Li Q., Stamatoyannopoulos G. 1994. Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood* **84**:1399-401.
Liu Q., Bungert J., Engel J.D. 1997. Mutation of gene-proximal regulatory elements disrupts human epsilon-, gamma-, and beta-globin expression in yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A* **94**:169-74.
Liu Q., Tanimoto K., Bungert J., Engel J.D. 1998. The A gamma-globin 3' element

provides no unique function(s) for human beta-globin locus gene regulation. *Proc Natl Acad Sci U S A* 95:9944-9.
Magram J., Chada K., Costantini F. 1985. Developmental regulation of a cloned adult beta-globin gene in transgenic mice. *Nature* 315:338-40.
Marin M., Karis A., Visser P., Grosveld F., Philipsen S. 1997. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* 89:619-28.
Martin D.I., Tsai S.F., Orkin S.H. 1989. Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature* 338:435-8.
Martin D.I., Fiering S., Groudine M. 1996. Regulation of beta-globin gene expression: straightening out the locus. *Curr Opin Genet Dev* 6:488-95.
Martin F., van Deursen J.M., Shivdasani R.A., Jackson C.W., Troutman A.G., Ney P.A. 1998. Erythroid maturation and globin gene expression in mice with combined deficiency of NF-E2 and nrf-2. *Blood* 91:3459-66.
Merika M., Orkin S.H. 1995. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol* 15:2437-47.
Miller I.J, Bieker J.J. 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* 13:2776-86.
Milot E., Strouboulis J., Trimborn T., Wijgerde M., de Boer E., Langeveld A., Tan-Un K., Vergeer W., Yannoutsos N., Grosveld F., Fraser P. 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* 87:105-14.
Moon A.M., Ley T.J. 1990. Conservation of the primary structure, organization, and function of the human and mouse beta-globin locus-activating regions. *Proc Natl Acad Sci U S A* 87:7693-7.
Mueller-Storm H.P., Sogo J.M., Schaffner W. 1989. An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell* 58:767-77.
Myers R.M., Tilly K., Maniatis T. 1986. Fine structure genetic analysis of a beta-globin promoter. *Science* 232:613-8.
Ney P.A., Sorrentino B.P., Lowrey C.H., Nienhuis A.W. 1990. Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucleic Acids Res* 18:6011-7.
Ney P.A., Sorrentino B.P., McDonagh K.T., Nienhuis A.W. 1990. Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev* 4:993-1006.
Nuez B., Michalovich D., Bygrave A., Ploemacher R., Grosveld F. 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* 375:316-8.
O'Neill D., Yang J., Erdjument-Bromage H., Bornschlegel K., Tempst P., Bank A. 1999. Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. *Proc Natl Acad Sci U S A* 96:349-54.
Orlando V., Jane E.P., Chinwalla V., Harte P.J., Paro R. 1998. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J* 17:5141-50.
Ouyang L., Chen X., Bieker J.J. 1998. Regulation of erythroid Kruppel-like factor (EKLF) transcriptional activity by phosphorylation of a protein kinase casein kinase II site

within its interaction domain. *J Biol Chem* **273**:23019-25.
Palena A., Blau A., Stamatoyannopoulos G., Anagnou N.P. 1994. Eastern European (delta beta) zero-thalassemia: molecular characterization of a novel 9.1-kb deletion resulting in high levels of fetal hemoglobin in the adult. *Blood* **83**:3738-45.
Papayannopoulou T., Brice M., Stamatoyannopoulos G. 1986. Analysis of human hemoglobin switching in MEL x human fetal erythroid cell hybrids. *Cell* **46**:469-76.
Perkins A.C., Sharpe A.H., Orkin S.H. 1995. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* **375**:318-22.
Peterson K.R., Clegg C.H., Huxley C., Josephson B.M., Haugen H.S., Furukawa T., Stamatoyannopoulos G. 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 U S A* **90**:7593-7.
Peterson K.R., Clegg C.H., Navas P.A., Norton E.J., Kimbrough T.G., Stamatoyannopoulos G. 1996. Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A* **93**:6605-9.
Pevny L., Lin C.S., D'Agati V., Simon M.C., Orkin S.H., Costantini F. 1995. Development of hematopoietic cells lacking transcription factor GATA-1. *Development* **121**:163-72.
Pirrotta V. 1998. Polycombing the genome: PcG, trxG, and chromatin silencing. *Cell* **93**:333-6.
Powell R., Gannon F. 1988. The leghaemoglobins. *Bioessays* **9**:117-21.
Ptashne M. 1988. How eukaryotic transcriptional activators work. *Nature* **335**:683-9.
Raich N., Enver T., Nakamoto B., Josephson B., Papayannopoulou T., Stamatoyannopoulos G. 1990. Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science* **250**:1147-9.
Raich N, Papayannopoulou T, Stamatoyannopoulos G, Enver T 1992. Demonstration of a human epsilon-globin gene silencer with studies in transgenic mice. *Blood* **79**:861-4.
Reik A., Telling A., Zitnik G., Cimbora D., Epner E., Groudine M. 1998. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol* **18**:5992-6000.
Shih D.M., Wall R.J., Shapiro S.G. 1990. Developmentally regulated and erythroid-specific expression of the human embryonic beta-globin gene in transgenic mice. *Nucleic Acids Res* **18**:5465-72.
Shivdasani R.A., Orkin S.H. 1995. Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc Natl Acad Sci U S A* **92**:8690-4.
Simon M.C., Pevny L., Wiles M.V., Keller G., Costantini F., Orkin S.H. 1992. Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. *Nat Genet* **1**:92-8.
Skow L.C., Burkhart B.A., Johnson F.M., Popp R.A., Popp D.M., Goldberg S.Z., Anderson W.F., Barnett L.B., Lewis S.E. 1983. A mouse model for beta-thalassemia. *Cell* **34**:1043-52.
Smith R.D., Yu J., Annunziato A., Seale R.L. 1984. beta-Globin gene family in murine erythroleukemia cells resides within two chromatin domains differing in higher order structure. *Biochemistry* **23**:2970-6.
Stamatayonnopoulos G. and A.W. Nienhuis 1996. Hemoglobin switching, the molecular basis of blood diseases. W.B. Saunders, Philidelphia, USA.

Stamatoyannopoulos J.A., Clegg C.H., Li Q. 1997. Sheltering of gamma-globin expression from position effects requires both an upstream locus control region and a regulatory element 3' to the A gamma-globin gene. *Mol Cell Biol* **17**:240-7.
Stanworth S.J., Roberts N.A., Sharpe J.A., Sloane-Stanley J.A., Wood W.G. 1995. Established epigenetic modifications determine the expression of developmentally regulated globin genes in somatic cell hybrids. *Mol Cell Biol* **15**:3969-78.
Starck J., Sarkar R., Romana M., Bhargava A., Scarpa A.L., Tanaka M., Chamberlain J.W., Weissman S.M., Forget B.G. 1994. Developmental regulation of human gamma- and beta-globin genes in the absence of the locus control region. *Blood* **84**:1656-65.
Steward T.A., Wagner E.F., Mintz B. 1982. Human beta-globin gene sequences injected into mouse eggs, retained in adults, and transmitted to progeny. *Science* **217**:1046-8.
Strouboulis J., Dillon N., Grosveld F. 1992a. Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev* **6**:1857-64.
Strouboulis J., Dillon N., Grosveld F. 1992b. Efficient joining of large DNA fragments for transgenesis. *Nucleic Acids Res* **20**:6109-10.
Talbot D., Collis P., Antoniou M., Vidal M., Grosveld F., Greaves D.R. 1989. A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* **338**:352-5.
Talbot D., Philipsen S., Fraser P., Grosveld F. 1990. Detailed analysis of the site 3 region of the human beta-globin dominant control region. *EMBO J* **9**:2169-77.
Tang D.C., Ebb D., Hardison R.C., Rodgers G.P. 1997. Restoration of the CCAAT box or insertion of the CACCC motif activates [corrected] delta-globin gene expression. *Blood* **90**:421-7.
Taramelli R., Kioussis D., Vanin E., Bartram K., Groffen J., Hurst J., Grosveld F.G. 1986. Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. *Nucleic Acids Res* **14**:7017-29.
Tewari R., Gillemans N., Wijgerde M., Nuez B., von Lindern M., Grosveld F., Philipsen S. 1998. Erythroid Kruppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region. *EMBO J* **17**:2334-41.
Townes T.M., Lingrel J.B., Chen H.Y., Brinster R.L., Palmiter R.D. 1985. Erythroid-specific expression of human beta-globin genes in transgenic mice. *EMBO J* **4**:1715-23.
Tsai S.F., Strauss E., Orkin S.H. 1991. Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes Dev* **5**:919-31.
Tsai F.Y., Keller G., Kuo F.C., Weiss M., Chen J., Rosenblatt M., Alt F.W., Orkin S.H. 1994. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**:221-6.
Tsang A.P., Visvader J.E., Turner C.A., Fujiwara Y., Yu C., Weiss M.J., Crossley M., Orkin S.H. 1997. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90**:109-19.
Tsang A.P., Fujiwara Y., Hom D.B., Orkin S.H. 1998. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* **12**:1176-88.
Tuan D., Solomon W., Li Q., London I.M 1985. The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A* **82**:6384-8.

Tuan D.Y., Solomon W.B., London I.M., Lee D.P. 1989. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A* **86**:2554-8.
Tuan D., Kong S., Hu K. 1992. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* **89**:11219-23.
Wall L., deBoer E., Grosveld F. 1988. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev* **2**:1089-100.
Van der Hoeven F., Zakany J., Duboule D. 1996. Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls. *Cell* **85**:1025-35.
Van der Ploeg L.H., Konings A., Oort M., Roos D., Bernini L., Flavell R.A. 1980. Gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature* **283**:637-42.
Webber A.L., Ingram R.S., Levorse J.M., Tilghman S.M. 1998. Location of enhancers is essential for the imprinting of H19 and Igf2 genes. *Nature* **391**:711-5.
Whitelaw E., Tsai S.F., Hogben P., Orkin S.H. 1990. Regulated expression of globin chains and the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse. *Mol Cell Biol* **10**:6596-606.
Wijgerde M., Grosveld F., Fraser P. 1995. Transcription complex stability and chromatin dynamics in vivo. *Nature* **377**:209-13.
Wijgerde M., Gribnau J., Trimborn T., Nuez B., Philipsen S., Grosveld F., Fraser P. 1996. The role of EKLF in human beta-globin gene competition. *Genes Dev* **10**:2894-902.
Wright S., Rosenthal A., Flavell R., Grosveld F. 1984. DNA sequences required for regulated expression of beta-globin genes in murine erythroleukemia cells. *Cell* **38**:265-73.
Zafarana G., Raguz S., Pruzina S., Grosveld F. and Meyer D. 1996. The regulation of beta globin gene expression: the analysis of HS5 in the LCR. Haemoglobin switching. G. Stamatoyannopoulos (ed.) . Intercept Ltd. Hampshire, UK.
Zhang W., Bieker J.J. 1998. Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. *Proc Natl Acad Sci U S A* **95**:9855-60.

Chapter 3.

The role of EKLF in human β-globin 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

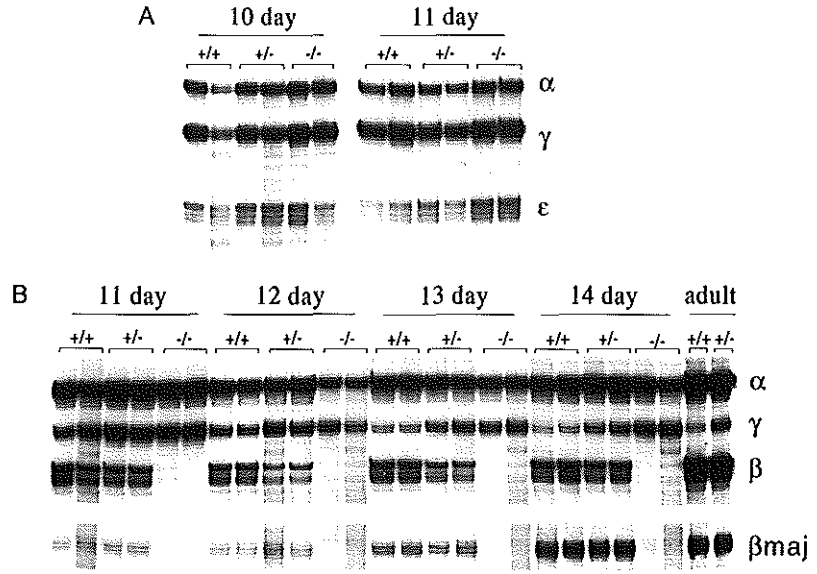


Figure 1. S1 nuclease protection assays. Probes for human ε, γ and β globin and mouse α and β-major globin were used as described in Materials and Methods. All mice analyzed were homozygous for a single copy of the entire human β-globin locus (line 72, Strouboulis et al. 1992) in either EKLF$^{+/+}$, 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

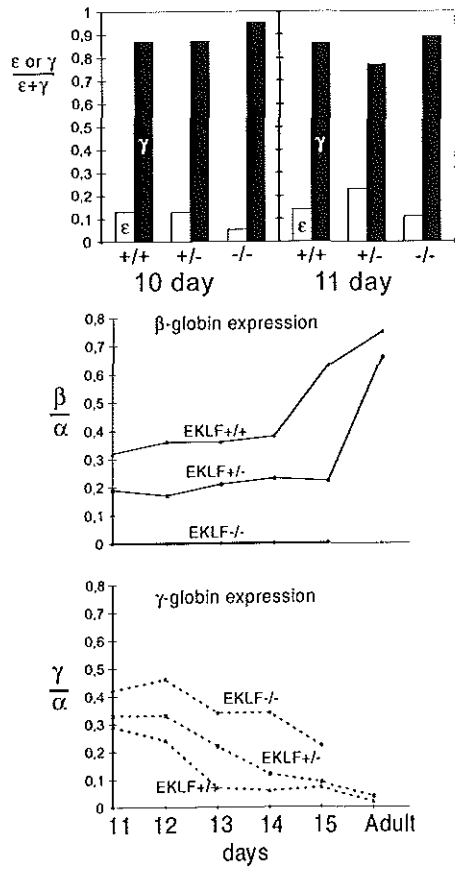


Figure 2. Expression pattern of the human globin transgenes in EKLF null background. Expression levels were quantitated using PhosphorImager analysis of S1 protection assays in Fig. 1, and plotted against the developmental age of the animals. (*A*) Relative expression levels of ε- and γ-globin mRNA in 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

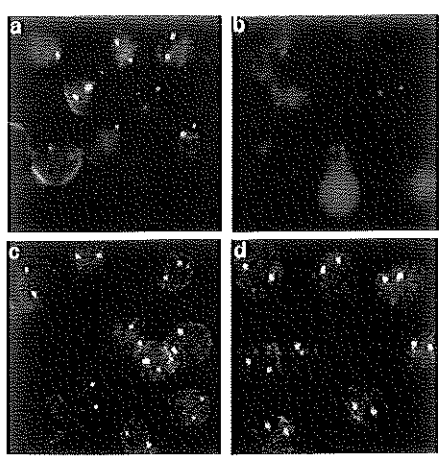


Figure 3. Primary transcript in situ hybridization. EKLF wild-type and knockout 13.5-day mouse fetal liver cells homozygous for the human β-globin locus were hybridized with gene-specific intron probes for (*A,B*) human γ (red) and β globin (green) or (*C,D*) mouse α globin (green) and human γ globin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

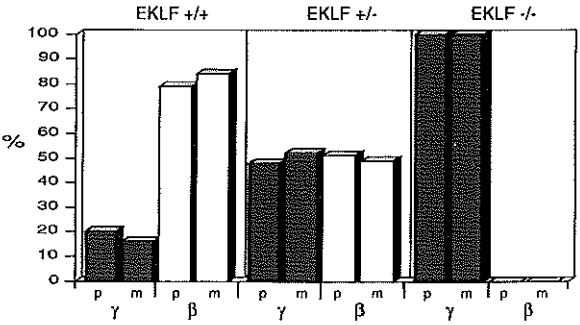


Figure 4. Correlation between transcriptional activity and steady-state mRNA levels. Quantities of steady-state mRNA (γ or β) at 13.5 days postconception from Figs. 1 and 2 are represented as a percentage of total human RNA (γ or β/γ + β). The percentages of transcriptionally active γ and β genes were 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.

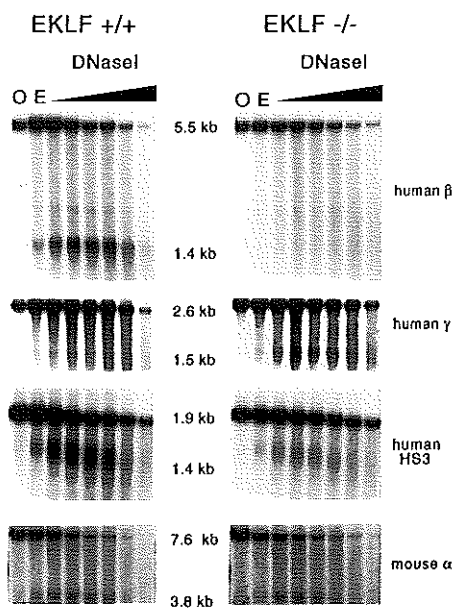


Figure 5. DNase I hypersensitive site analysis of 13.5-day fetal liver cells. Shown are DNase I hypersensitivity analyses in isolated nuclei from fetal livers of mice homozygous for the human β-globin locus in EKLF$^{+/+}$ and EKLF$^{-/-}$ backgrounds. The probes used are described in Materials and Methods 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. Stamatoyannopoulos, 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.

Chromatin interaction mechanism of transcriptional control in vivo.

Gribnau, J., de Boer, E., Trimborn, T., Wijgerde, M., Milot, E., Grosveld, F. and Fraser, P.
EMBO J. 17, 6020-6027, (1998).

Joost Gribnau, Ernie de Boer, Tolleiv Trimborn, Mark Wijgerde, Eric Milot, Frank Grosveld and Peter Fraser[1]

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

[1]Corresponding author
e-mail: fraser@ch1.fgg.eur.nl

We have used a kinetic analysis to distinguish possible mechanisms of activation of transcription of the different genes in the human β globin locus. Based on *in situ* studies at the single-cell level we have previously suggested a dynamic mechanism of single genes alternately interacting with the locus control region (LCR) to activate transcription. However, those steady-state experiments did not allow a direct measurement of the dynamics of the mechanism and the presence of loci with *in situ* primary transcript signals from two β-like genes *in cis* has left open the possibility that multiple genes in the locus could initiate transcription simultaneously. Kinetic assays involving removal of a block to transcription elongation in conjunction with RNA FISH 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.

Keywords: α globin/β globin/locus control region/single gene activation/transcription

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; Mueller-Storm *et al.*, 1989; Bickel and Pirotta, 1990; 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 from 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 (Tuan *et al.*, 1992; Wijgerde *et al.*, 1995; Martin *et al.*, 1996; 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 (Figure 1A). The expression of all of these genes is dependent on the presence of the LCR which is located 15 kb upstream of 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 weeks 6 and 10 of gestation. Expression of the γ genes predominates during the fetal liver stage. In the later 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 entire human β locus is incorporated 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 suppressed autonomously through sequences directly flanking the genes (Raich *et al.*, 1990; Dillon and Grosveld, 1991). 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 accessibility 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

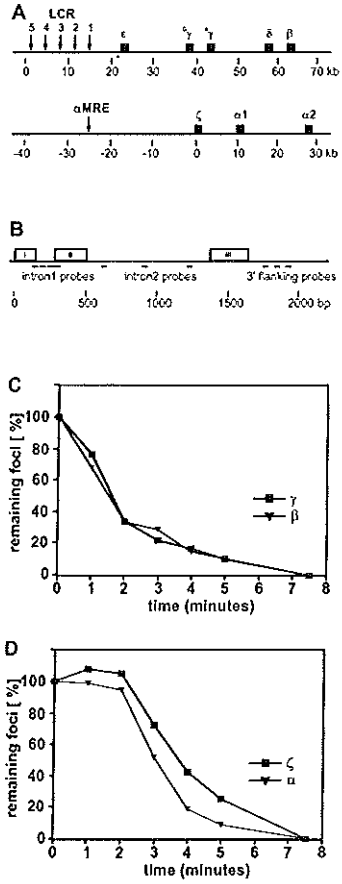


Fig. 1. Lifetime of the human β and mouse α globin primary transcript signals *in situ*. (**A**) Schematic diagrams of the human β globin locus and mouse α globin locus. Black boxes represent genes and vertical arrows represent the hypersensitive sites of the β globin LCR and the α globin major regulatory element (αMRE). (**B**) Intron/exon structure of a β-like globin gene with localization of probes used for *in situ* hybridizations. (**C**) Decay of the γ and β primary transcripts *in situ* signals in day 11.5 fetal liver cells treated with 5 μg/ml actinomycin-D. The percentage of remaining intron 2 signals are shown as a function of time. (**D**) As (**C**) for the mouse ζ and α genes in day 10.5 embryonic blood cells.

the scanning or the accessibility model of transcription initiation (Martin *et al.*, 1996). Thus a crucial difference between the looping model and the others is single versus multiple gene activation at any moment in a single locus. We have therefore used a novel kinetic analysis utilizing inhibition of transcription elongation and release in conjunction with RNA fluorescence *in situ* hybridization (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 primary transcript 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 rapidly cleaved from the primary transcript and degraded. We used actinomycin-D to measure the detection lifetime of the γ and β primary transcript signals at days 11.5 (Figure 1C) and 12.5 (not shown), and ζ and α primary transcript signals at day 10.5 (Figure 1D) of development. On both day 11.5 and 12.5 the intron signals of the γ and β genes, as well as ζ and α intron signals at day 10.5, disappear below detection level 7.5 min after the addition of actinomycin-D. Interestingly the intron signals of the ζ and α genes do not decrease immediately when compared with 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-β-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 effect initiation of transcription (Fraser *et al.*, 1978; Marshall and Price, 1992) but prematurely aborts elongating transcripts ~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 C-terminal domain (CTD) of pol II (Marshall *et al.*, 1996; Peng *et al.*, 1998). The effect of DRB treatment on globin gene transcription in mouse erythroleukemia cells has been reported previously (Tweeten and Molloy, 1981). The results demonstrate that DRB causes premature termination without affecting initiation of transcription. 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 (Figures 1B and 2). *In situ* signals with probes that hybridize to intron 1, located in the first 300 bases of the β globin primary transcript, are still visible in 85% of the erythroid cells after 15 min of DRB treatment (Figure 2A and C) when compared with the untreated control. Probes that hybridize to intron 2, 600–1200 bases 3′ of the initiation site (Figure 1B), have completely disappeared after 7.5 min of DRB treatment (Figure 2B and C). The fact that intron 1 signals are not affected by DRB confirms earlier reports which indicated that the process of transcription initiation is not disturbed and 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

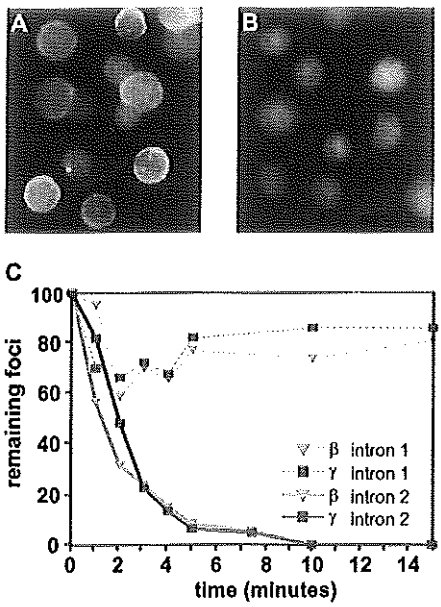


Fig. 2. The effect of DRB on transcriptional elongation *in vivo*. (A) *In situ* hybridization with intron-1-specific probes on homozygous transgenic day 11.5 fetal liver cells treated for 15 min with 100 μM DRB. γ globin signals are shown in red (Texas red), and β globin signals are green (FITC). (B) *In situ* hybridization with intron-2-specific probes on homozygous transgenic 11.5 day fetal liver cells treated for 7 min with 100 μM DRB. γ globin signals are red (Texas red), and β globin signals are green (FITC). (C) Decay of γ and β intron 1 and intron 2 *in situ* signals plotted as a function of time after the addition of DRB.

is not disturbed by the addition of DRB. The results also show 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 (Figure 1C). The same is found for the α-like 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 ~85% of the human globin loci (Wijgerde *et al.*, 1995). A small percentage of loci (~15%, e.g. Figure 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 according to the scanning and (on a random basis) accessibility models, then release of the DRB block should result in the reappearance of double signals (using intron 2 probes) at the same rate as single signals. A lag in the reappearance of double signals compared with single gene signals would be indicative of alternating single gene initiation. This model predicts that double signals would result from the overlap between decaying primary transcripts (7.5 min, see above) from a recently active gene and the nascent transcription of an active gene.

We treated transgenic 11.5 day homozygous fetal liver cells with DRB for 15 min to block elongation of globin primary transcripts (Figure 3). The cells were then released from the DRB block by washing with phosphate-buffered saline (PBS), and aliquots of cells were fixed onto slides at various intervals and prepared for *in situ* hybridization as described previously (Wijgerde *et al.*, 1995). The cells were probed with intron 2-specific probes for human γ (probes detect both $^{G}\gamma$ and $^{A}\gamma$ primary transcripts) and β globin primary transcripts. No transcription signals are visible at the zero time point immediately after the wash (Figure 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 and in the non-DRB treated controls. Single (γ or β) and double (γ and β *in cis*) gene signals were counted and the averaged results of four separate experiments from two transgenic lines, line 72 and line 2 (Strouboulis *et al.*, 1992), are presented in Figure 4. The results show that the reappearance of double gene signals significantly lags behind the reappearance of single gene signals in the population. A lag in the reappearance of double signals is also consistently observed when the two γ genes are compared with gene-specific probes for $^{G}\gamma$ and $^{A}\gamma$ (data not shown) and for the genes in the endogenous mouse β-globin locus (T.Trimborn, J.Gribnau, F.Grosveld, M.Wijgerde and P.Fraser, submitted).

To determine whether there is an intrinsic bias against the reappearance 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 (Figure 4; loci trans) closely approximates to 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 appearance of two signals *in trans*. We therefore conclude that the two allelic globin loci behave independently of each other in terms of transcription (showing a stochastic reappearance of 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).

Weak competition versus strong competition

The γ genes are being expressed in embryonic cells in conjunction with ε and in the early fetal liver cells in competition with β. This provides a unique opportunity to investigate the interdependence of genes within a locus by examining the effects of weak versus strong gene competition during development independent of a DRB treatment. The ε gene, which is a weak competitor (Wijgerde *et al.*, 1995), increases in expression from 9.5 to 11.5 days in the embryonic blood and γ gene expression decreases reciprocally (Strouboulis *et al.*, 1992; Table I). The ε-globin gene is not expressed in the early fetal liver erythroid cells, but γ expression is further decreased due

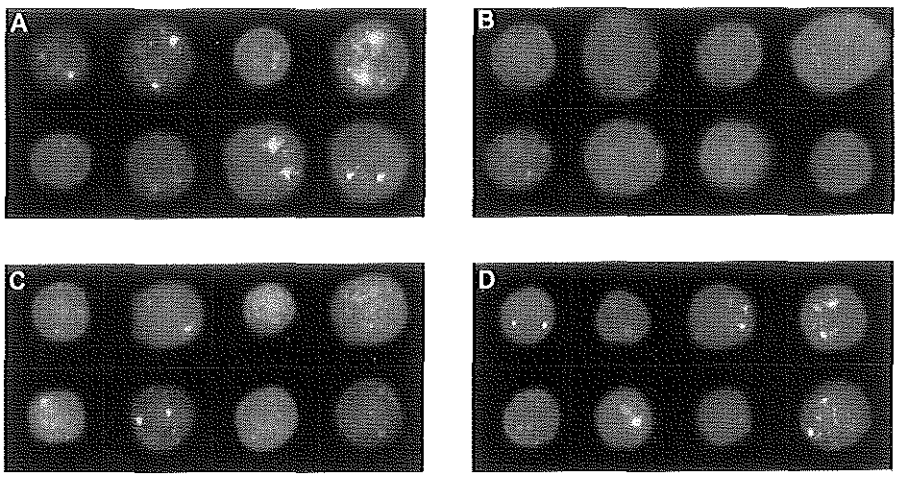


Fig. 3. Reversible inhibition of transcription elongation on homozygous 11.5 day transgenic fetal liver cells. (A–D) Primary transcript *in situ* hybridizations using γ and β intron 2 probes; γ signals are red (Texas red) and β signals are green (FITC). (**A**) γ and β primary transcript signals prior to DRB treatment. (**B**) γ and β primary transcript signals after 15 min treatment with 100 μM DRB. (**C**) Five minutes after release of the transcriptional elongation block by washing out DRB. Note most cells show single γ or β transcription signals *in cis*, while many have more than one signal *in trans*. (**D**) Twenty minutes after release from the elongation block, the distribution of loci having single γ or β and double signals (γ and β *in cis*) are back to control levels. Representative cells are shown for each time point.

to strong competition from the highly expressed β gene (Wijgerde *et al.*, 1995, 1996). If co-initiation of transcription occurs one would predict that the percentage of $^{G}\gamma\text{–}^{A}\gamma$ double signals in the γ-expressing cells would change very little. In contrast, if alternating transcription of the γ genes and the other globin genes (ε or β) occurs then addition of a third gene into the competition should affect the percentage of double γ signals due to the resulting three-way alternation. When the percentage of $^{G}\gamma\text{–}^{A}\gamma$ double signals is measured in the γ-expressing cells during development, it changes from 83% at day 9.5 when ε expression is low, to 56% at day 11.5 in embryonic blood when ε expression is at its maximum (Table 1). In the fetal liver where the γ genes are expressed with high levels of β expression, the percentage of double $^{G}\gamma\text{–}^{A}\gamma$ signals decreases further to 29% (Table 1). Thus when γ gene expression is accompanied by relatively low-level expression of the ε gene in embryonic cells or high-level expression of the β gene in fetal liver cells, the percentage of $^{G}\gamma\text{–}^{A}\gamma$ double signals decreases accordingly. This result is most easily explained by an alternating single gene mechanism and is difficult to explain if one assumes that the γ genes are co-initiated.

Transcription initiation of the α-like globin genes

The DRB analysis above does not exclude the possibility that treatment with the drug could somehow artifactually cause a delay in the reappearance of double signals *in cis*, even though this is clearly not the case for the reappearance of double signals *in trans* (Figure 4). There is indirect evidence that the α-globin genes may be regulated differently from the β-like genes (Craddock *et al.*, 1995). Instead of the five hypersensitive regions present in the β globin LCR, only a single hypersensitive site has been identified

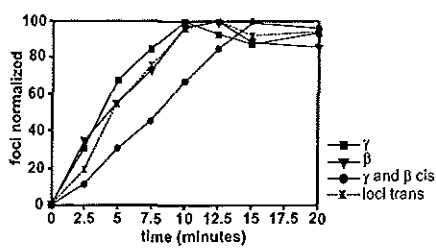


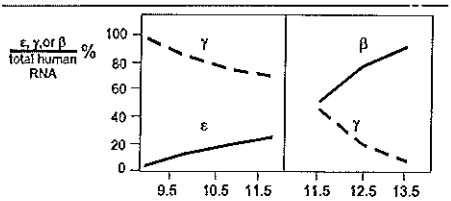
Fig. 4. Kinetics of reappearance of single versus double gene signals. Primary transcript *in situ* signals for γ and β (see Figure 3) were scored after release of the DRB block to transcription elongation in 11.5 day fetal liver cells from lines 72 and 2 (Strouboulis *et al.*, 1992) and plotted versus time. The curves were normalized to their maximum values and the average of four experiments is shown. All experiments show the early reappearance of single γ or β signals and a clear lag in reappearance of double (γ and β *in cis*) signals. Cells in which both homologues display signals for either γ or β are plotted as loci trans showing the independent reappearance of two signals in the same cell.

40 or 26 kb upstream of the human and mouse α genes, respectively (Gourdon *et al.*, 1994, 1995). When this site is present as part of the human α locus in transgenic mice, the expression of the α genes is supp essed as development proceeds (Sharpe *et al.*, 1993; Gcurdon *et al.*, 1994), indicating that additional sequences are required for full expression.

We examined the primary transcription of the mouse α-like genes in embryonic red cells in which all three α genes are expressed. Primary transcript *in situ* analysis shows a high percentage of double ζ and α signals *in cis* (80%) versus single ζ or α signals (Figure 5A), while the

lifetime of the signal is similar to that observed for the β signals (Figure 1D) in the presence of actinomycin-D as well as DRB (data not shown). This could mean that the genes are co-initiated in many loci or that frequently alternating initiation takes place. If alternating initiation occurs 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 (Figure 5B–D). Plotting the reappearance of the signals as a function of time (Figure 6) 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 cannot be excluded. Importantly, the result shows that the treatment with DRB does not artificially result in a lag in the appearance of double signals *in cis*.

Table 1. Developmental expression and transcription of the human γ genes



	% 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

The top panel shows the relative expression of human ε, γ and β genes during development in transgenic mice (Strouboulis *et al.*, 1994). Shown below are the percentages of loci with single $^{G}\gamma$ or $^{A}\gamma$ signals and double $^{G}\gamma$–$^{A}\gamma$ primary transcript *in situ* signals during development. The bottom right shows the percentage of $^{G}\gamma$ and $^{A}\gamma$ mRNA of total γ mRNA as determined by primer extension (data not shown). Note that changes in steady-state RNA levels (accumulated mRNA) lag behind changes in transcription as detected by primary transcript *in situ* hybridization.

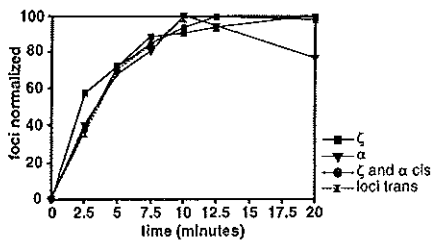


Fig. 6. Kinetics of reappearance of single versus double ζ and α primary transcript signals. *In situ* signals as shown in Figure 5 were scored after release of the transcriptional elongation block in 10.5 day embryonic blood cells and normalized to their maximum values for single (ζ or α) and double (ζ and α *in cis*) signals. Cells in which both homologues display signals for either ζ or α are plotted as loci trans showing the independent reappearance of two signals in the same cell.

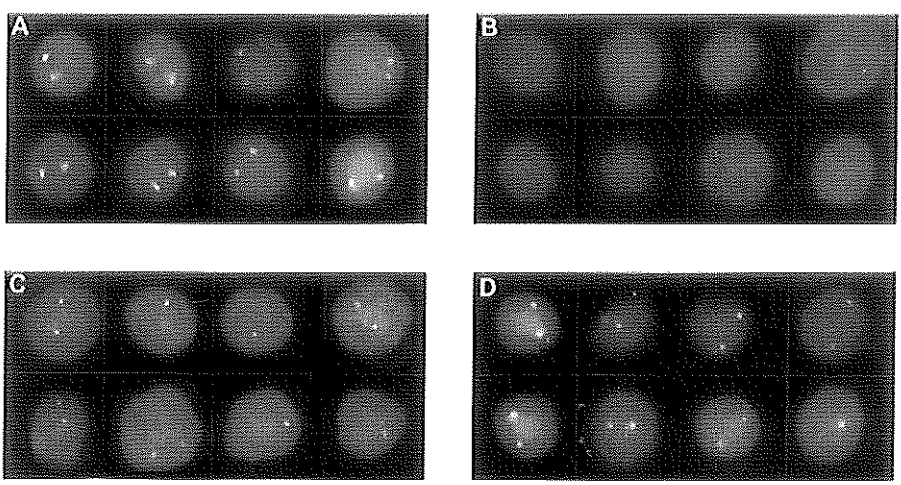


Fig. 5. Reversible inhibition of transcription elongation on day 10.5 blood cells. (A–D) Primary transcript *in situ* hybridizations using mouse ζ and α intron probes; ζ signals are shown in red (Texas red) and α signals in green (FITC). (A) ζ and α signals prior to DRB treatment. (B) Fifteen minutes after treatment with 100 μM DRB. (C) Five minutes after release of the transcriptional elongation block by washing out DRB. Note that many loci show double ζ- and α-globin transcription signals *in cis*. (D) Twenty minutes after release of elongation block, the distribution of loci having single ζ or α and double signals (ζ and α *in cis*) are back to control levels. Representative cells are shown for each time point.

Discussion

The results obtained previously with primary transcript *in situ* analysis suggested that 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. 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). These data, taken together with the primary transcript *in situ* results which revealed the presence of a minority of loci with double primary transcription signals *in cis*, were interpreted to mean that the γ and β globin genes within a single locus were or could be co-initiated (Bresnick and Felsenfeld, 1994; Furukawa *et al.*, 1994; Martin *et al.*, 1996).

Three models have been proposed to explain transcriptional regulation by distant regulatory sequences. The accessibility model proposes that 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). The scanning model suggests that 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). Finally the looping model postulates that the distant regulatory sequences interact directly with the gene to initiate transcription (Ptashne, 1988; Wijgerde *et al.*, 1995 and references therein). Both the accessibility and the scanning models allow co-initiation of multiple genes *in cis*, whereas the looping model predicts initiation of only one gene at a time. It is therefore important in understanding the process of transcriptional initiation *in vivo* involving distant regulatory sequences to distinguish whether alternating single or multiple co-initiation takes place.

We have shown that in the presence of DRB *in vivo*, initiation of transcription is unaffected whereas transcriptional elongation is reversibly blocked. Inhibition studies with both DRB and actinomycin-D show that the globin primary transcript signals have a maximum lifetime of ~7 min, confirming that we are detecting actively transcribed genes or genes transcribed within the last 7 min. By reversibly inhibiting elongation with DRB for 15 min we have allowed intron 2 primary transcript signals to decay via splicing and degradation. Removal of the block theoretically permits the immediate detection of only actively transcribed genes. Obviously, not all loci cross the detection threshold simultaneously, introducing a small but measurable degree of asynchrony to the experiment, and therefore allowing the possibility of the reappearance of some double signals in the earliest time point. By analyzing large populations of cells at various time points after removal of the elongation block we have been able to demonstrate a clear and consistent difference in the amount of time required to detect single versus double signals. If we compare the $t_{1/2max}$ (the time required to re-establish 50% of the maximum number of single or double foci) for single (γ or β) and double (γ and β) signals, we find 4 and 8 min, respectively. This shows that on average double signals take twice as long to reappear as single signals, indicating that double signals result from sequential periods of initiation of single genes. The fact that the same result is obtained for the human β globin locus in two different transgenic lines and that it is essentially the same as observed for the endogenous murine β-globin locus (T.Trimborn, J.Gribnau, F.Grosveld, M.Wijgerde and P.Fraser, submitted) shows that the results are not due to a 'transgenesis' artefact. Interestingly, although the human β-globin locus in line 72 is integrated in a euchromatic environment, while in line 2 it is integrated in a heterochromatic environment close to the centromere, this appears to make no difference in terms of mechanism.

In contrast, in the α locus the $t_{1/2max}$ of the single and double signals are essentially the same (~3 min) suggesting that the α genes are either co-initiated or that they alternate at a frequency which is shorter than the time required to reach the detection threshold. There is indirect evidence for competition in the mouse α locus which may be indicative of an alternating transcriptional mechanism. Insertional mutation of the ζ gene with a PGK-Neo cassette results in decreased expression of the α genes in definitive erythroid cells when ζ would normally be silenced (Leder *et al.*, 1997). Although the DRB results alone do not allow us to make firm conclusions regarding the mechanism of multiple α gene expression they do show a clear difference compared with the γ and β genes and exclude an artificial lag of reappearing double signals *in cis* as a consequence of the DRB treatment.

Restriction digestion of sites in the $^{G}\gamma$ and $^{A}\gamma$ promoter regions in isolated K562 nuclei indicated that both promoters were accessible to digestion *in cis* in ~50% of the loci (Bresnick and Felsenfeld, 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 been established formally, it was suggested that this was an indication of co-initiation of transcription of the genes. Our DRB results in conjunction with the developmental transcription analysis indicate that the two γ genes are alternately transcribed. The high percentage of double $^{G}\gamma$–$^{A}\gamma$ signals in comparison with the low percentage of double γ–β signals suggests that alternations occur at a higher frequency between the two γ genes than between the γ genes and the β gene.

Thus the data presented show that the double signals in the β locus are not generated as predicted by a co-initiation mechanism, but represent a transition between alternating transcriptional periods of single genes. A number of additional *in vivo* observations support this conclusion. A single regulatory region (the LCR) is required by all of the genes in the β globin locus for activation (reviewed in Fraser *et al.*, 1998) and the genes compete with each other for this function (Giglioni *et al.*, 1984; Behringer *et al.*, 1990; Enver *et al.*, 1990; Hanscombe *et al.*, 1991; Dillon *et al.*, 1997), with LCR proximal genes having a competitive advantage over distal genes (Hanscombe *et al.*, 1991; Peterson and Stamatoyannopoulos, 1993; Dillon *et al.*, 1997). During the period of switching from γ to β expression nearly all erythroid cells have both γ and β mRNA in the cytoplasm, yet the overwhelming majority of loci have only γ or β transcription signals (Wijgerde *et al.*, 1995). In addition, individual loci within the same cell can respond differently

to the same *trans*-acting factor environment with γ transcription on one homologue and β on the other. The balance of expression between the γ and β genes can be tipped in either direction by mutations in the γ promoter which prevent normal γ gene silencing or alterations in the level of EKLF, which is required for β gene transcription (Wijgerde *et al.*, 1996; Tewari *et al.*, 1998). In each case modulation of the expression of one gene leads to reciprocal changes in the expression of the other. A dynamic interaction between the LCR and the genes via looping explains all of the basic properties. 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 competitive advantage of a gene is the result of increased frequency of LCR–gene interactions, which are dependent on distance (see Dillon *et al.*, 1997).

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 (Dunaway and Dröge, 1989; Mueller-Storm *et al.*, 1989; Bickel and Pirotta, 1990). It implies that direct chromatin interactions between the LCR and a single gene are required for initiation of transcription and suggests that continued loading of polymerases or re-initiation 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 suggest 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.

Materials and 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 from the indicated developmental time points were dissected out in PBS. Fetal livers were gently disrupted by repeated pipetting. Actinomycin-D was used at a final concentration of 5 μg/ml. DRB (Sigma) was added to cell suspensions to a final concentration of 100 μM and incubated at 37°C for 15 min. Five volumes of ice cold PBS were added and the cells were immediately pelleted by centrifugation for 2 min at 1500 r.p.m. in an Eppendorf centrifuge. Cells were washed twice more with 1.5 ml ice-cold PBS and resuspended in 250 μl 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.

Probe sequences and in situ hybridization analysis

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

Human β intron 1 probes:

5′-CTGTCTCCACATGCCCAGTTTCTATGGTCTCCTTAAACCTG-TCTTGTAA-3′
5′-GGGTGGGAAAATAGACCAAAGGCAGAGAGAGTCAGTGCC-TATCAGAAAC-3′
5′-AGGGCAGTAACGGCAGACTCTCCTCAGGAGTCAGGT-3′
5′-ATAACAGCATCAGGAGGGACAGATCCCCAAAGGACTCA-3′

Human β intron 2 probes:

5′-TTCCACACTGATGCAATCATTCGTCGTTTCCCATTCTAACT-GTACCCT-3′
5′-CTGATTTGGTCAATATGTGTACACAATTAAAACATTACACT-TTAACCCA-3′
5′-GGTAGCTGGATTGTAGCTGCTATAGCAATATGAAACCTCTT-ACATCAGT3′

Human γ intron1probes:

5′-AGGCACAGGGTCCTTCCTTCCCTCCCTTGTCCTGGTCAC-3′
5′-TGACAAGAACAGTTTGACAGTCAGAAGGTGCCACAATCCT-GAGAAGCGA-3′
5′-AGGCTTGTGATAGTAGCCTTGTCCTCCTCTGTGAAATGA-CCCA-3′
5′-AGAGCCTACCTTCCCAGGGTTTCTCCTCCAGCATCTTCCA-CATT-3′

Human γ intron 2 probes:

5′-GCAGTTTCTTCACTCCCAACCCCAGATCTTCAAACAGCTCA-CACCCGC-3′
5′-CCTTCTGCCTGCATCTTTTTAACGACCAACTTGTCCTGCCT-CCAGAAG-3′
5′-ACAGAGCTGACTTTCAAAATCTACCCAGCCCAAATGTTTC-AATTGTCC-3′

Human $^{A}\gamma$ 3′ flanking region probes:

5′-TCATATAAAAATAAATGAGGAGCATGCACACACCACAAACA-CAAACAGGC-3′
5′-CAGAACTCCCGTGTACAAGTGTCTTTACTGCTTTTAT-3′
5′-TTCATTAAGAACCATCCTTGCTACTAGCTGCAATCAATCCA-GCCCCCA-3′
5′-ATTTCACTTTCTTAGGCATCCACAAGGCTGTGAAAAGCTAA-GTGCCAT-3′

Human $^{G}\gamma$ 3′ flanking region probes:

5′-AAAAAAGTGTGGAGTGTGCACATGACACAAACACACATAG-CCATGTATAA-3′
5′-TGCAGACGCTCCCATGTATAAGTTTCTTTATTGCCTAGTTCT-TTTATTTG-3′
5′-ACGTAAACAAAAAAGTGTGGAGTGGCACATGACACAAACA-CACATAG-3′
5′-GCAGACGCTCCCATGTATAAGTTCTTTATTGCCTAGTTCTTTT-ATTT-3′

Mouse α intron probes:

5′-CACAGAAAAGCATAGTTAGAAGCGCCCACTGAGCGAGTGC-CAGGTCC-3′
5′-AGCCCTTCCTAGGGGCCCAGATGCCGCCTGCCAGGTCCC-3′
5′-GCTCCCCTTCCTGGGACCACTATGTCCCTGCCTTGGGCACG-AGGACCC-3′

Mouse ζ intron probes:

5′-CCTTCTCAGTGGCTTCTCCTCACAACTGCTCTTTGTCACTTC-TGTCTC-3′
5′-ATGGAAGACTCTGGTGAGCTCTGGAATGCCAGCCCACCTC-CTTTAGTA-3′
5′-ACAACCCCAAGAGTGATGTTACTATTGCTGTTGCACAAGGG-TCTACA-3′
5′-AAGGGGATTTGATGCCTCCAGCCCCAATGGCACCCATGCCT-GCGCTCG-3′

The two γ genes are highly homologous in both intron and exon sequence making the use of gene-specific intron probes extremely difficult. Transcriptional termination of γ gene transcription is known to occur 1–2 kb downstream of the polyadenylation site (Ashe *et al.*, 1997). We therefore used probes (3′ flanking region probes) which hybridize to regions 300–500 bp downstream of the polyadenylation sites of the γ genes to detect the $^{G}\gamma$ and $^{A}\gamma$ gene primary transcripts separately.

Cells were fixed onto poly-L-lysine coated slides in 4% formaldehyde/5% acetic acid for 18 min at room temperature. The cells were subsequently washed three times for 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 0.01 M HCl, followed by a short wash in water and a 5 min fixation in 3.7% formaldehyde at room temperature. 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 24 × 24 mm coverslip) and incubated at 37°C in a humidified chamber for 12 h. The hybridization mixture contained 1 ng/μl of each oligonucleotide probe haptenized with either digoxygenin or biotin side chains in the middle and on the 5′ and 3′ ends of the oligonucleotide (Eurogentec, Belgium) in 25% formamide, 2× SSC, 200 μg/ml salmon sperm DNA, 5× Denhardt's, 1 mM EDTA and 50 mM sodium phosphate pH 7.0. The coverslip was removed by dipping in 2× SSC and the cells were washed three times for 10 min 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 at room

temperature. 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 epifluorescence microscopy and photographs recorded with an CCD camera.

The graphical results represent the average of three separate experiments. In all cases >1000 cells were counted per data point using a dictaphone to record the results.

Acknowledgements

We wish to thank Dr M.Antoniou for advice on DRB and actinomycin-D, Drs T.McMorrow and S.Philipsen 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.

References

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

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.

Bickel,S. and Pirotta,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.

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

Craddock,C.F., Vyas,P., Sharpe,J.A., Ayyub,H., Wood,W.G. and Higgs,D.R. (1995) Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments. *EMBO J.*, **14**, 1718–1726.

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

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

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

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

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

Foley,K.P. and Engel,J.D. (1992) 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.

Fraser,N.W., Sehgal,P.B. and Darnell,J.E. (1978) DRB-induced premature termination of late adenovirus transcription. *Nature*, **272**, 590–593.

Fraser,P., Gribnau,J. and Trimborn,T. (1998) Mechanisms of developmental regulation in globin loci. *Curr. Opin. Hematol.*, **3**, 139–144.

Furukawa,T., Zitnik,G., Leppig,K., Papayannopoulou,T. and Stamatoyannopoulos,G. (1994) 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**, 1412–1419.

Giglioni,B., Casini,C., Mantovani,R., Merli,S., Comi,P., Ottolenghi,S., Saglio,G., Camaschella,C. and Mazza,U. (1984) A molecular study of a family with Greek hereditary persistance of haemoglobin and β-thalassaemia. *EMBO J.*, **11**, 2641–2645.

Gourdon,G., Sharpe,J.A., Wells,D., Wood,W.G. and Higgs,D.R. (1994) 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. *Nucleic Acids Res.*, **22**, 4139–4147.

Gourdon,G., Sharpe,J.A., Higgs,D.R. and Wood,W.G. (1995) The mouse alpha-globin locus regulatory elements. *Blood*, **86**, 766–775.

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

Grosveld,F., Dillon and Higgs,D. (1993) The regulation of human globin gene expression. In Higgs,D.R. and Weatherall,D.J. (eds), *Bailliere's Clinical Haematology: The Haemoglobinopathies*. Bailliere Tindall, London, UK, vol. 6/No. 1, pp. 31–66.

Groudine,M. and Weintraub,H. (1982) Propagation of globin DNAase 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.

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.

Leder,A., Daugherty,C., Whitney,B. and Leder,P. (1997) Mouse ζ- and α-globin genes: embryonic survival, α-thalassemia and genetic background effects. *Blood*, **90**, 1275–1282.

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

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

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

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

Mueller-Storm,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.

Peng,J., Zhu,Y., Milton,J.T. and Price,D.H. (1998) Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev.*, **12**, 755–762.

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

Peterson,K., Clegg,C.H., Huxley,C., Josephson,B.M., Haughen,H.S., Furukawa,T. and Stamatoyannopoulos,G. (1993) Transgenic mice containing a 248kb YAC carrying the human β-globin locus display proper developmental control of human globin genes. *Proc. Natl Acad. Sci. USA*, **90**, 7593–7597.

Ptashne,M. (1988) How eukaryotic transcriptional activators work. *Nature*, **335**, 683–685.

Raich, N., Enver,T., Nakamoto,B., Josephson,B., Papayannopoulou,T. and Stamatoyannopoulos (1990) 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. and Wood,W.G. (1993) Analysis of the human alpha-globin gene cluster in transgenic mice. *Proc. Natl Acad. Sci. USA*, **90**, 11262–11266.

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.

Tewari,R., Gillemans,N., Wijgerde,M., Nuez,B., von Lindern,M., Grosveld,F. and Philipsen,S. (1998) Erythroid Kruppel-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.*, **17**, 2334–2341.

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

Tweeten,K.A. and Molloy,G.R. (1981) Induction of premature termination of transcription of the mouse β globin gene by DRB. *Nucleic Acids Res.*, **9**, 3307–3319.

Wijgerde,M., Grosveld,F. and Fraser,P. (1995) Transcription complex stability and chromatin dynamics. *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 May 21, 1998; revised July 24, 1998; accepted August 10, 1998

Chapter 5.

Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci.

Trimborn, T., Gribnau, J., Grosveld, F.and Fraser P.
Genes & Development 13, 112-124, (1999).

Mechanisms of developmental control of transcription in the murine α- and β-globin loci

Tolleiv Trimborn, Joost Gribnau, Frank Grosveld, and Peter Fraser[1]

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

We have characterized mRNA expression and transcription of the mouse α- and *β-globin* loci during development. S1 nuclease and primary transcript in situ hybridization analyses demonstrate that all seven murine globin genes (*ζ, α1, α2, εy, βH1, βmaj,* and *βmin*) are transcribed during primitive erythropoiesis, however transcription of the *ζ, εy,* and *βH1* genes is restricted to the primitive erythroid lineage. Transcription of the *βmaj* and *βmin* genes in primitive cells is EKLF-dependent demonstrating EKLF activity in embryonic red cells. Novel kinetic analyses suggest that multigene expression in the β locus occurs via alternating single-gene transcription whereas coinitiation cannot be ruled out in the α locus. Transcriptional activation of the individual murine β genes in primitive cells correlates inversely with their distance from the locus control region, in contrast with the human β locus in which the adult genes are only activated in definitive erythroid cells. The results suggest that the multigene expression mechanism of alternating transcription is evolutionarily conserved between mouse and human β globin loci but that the timing of activation of the adult genes is altered, indicating important fundamental differences in globin gene switching.

[*Key Words*: Globin genes; transcription; locus control region; gene competition; mechanism of transcriptional control; in vivo]

Received April 20, 1998; revised version accepted November 7, 1998.

The murine α- and β-globin loci are multigene clusters located on chromosomes 11 and 7, respectively (Fig. 1). They are highly homologous to their human counterparts in organizational structure and function and represent paradigms for the study of developmental gene regulation. The α locus consists of three genes (*ζ, α1,* and *α2*) that are dependent for expression on the major regulatory element (αMRE) which appears as an erythroid-specific DNase I-hypersensitive site ~26-kb upstream of the *ζ* gene (Gourdon et al. 1995). The murine β locus consists of four functional genes (*εy, βH1, β-major* (*βmaj*), and *β-minor* (*βmin*)) that are controlled by the locus control region (LCR), a series of five DNase I-hypersensitive sites in erythroid chromatin located 5–25 kb upstream of the *εy* gene (Moon and Ley 1990; Hug et al. 1992; Jimenez et al. 1992; Fiering et al. 1995). The genes are arranged in the order of their developmental expression, as are their human homologs. Embryonic yolk sac-derived erythroid cells coexpress high levels of both *ζ*- and α-globin mRNA (Leder et al. 1992) and primarily *εy* and *βH1* β-like globin mRNA with small amounts of *βmaj* and *βmin* (Brotherton et al. 1979; Chui et al. 1979; Wawrzyniak and Popp 1987; Whitelaw et al. 1990). At 11.5 days of gestation the major site of erythropoiesis in the developing embryo switches from the yolk sac to the fetal liver. This switch in site is coincident with a change to definitive gene expression in both the α and β clusters leading to predominant expression of the *α1* and *α2* genes and the *βmaj* and *βmin* genes. Although the small amount of *βmaj* and *βmin* expression in embryonic cells appears to be genuine and not caused by maternal contamination (Wawrzyniak and Popp 1987) it is unclear whether the embryonic genes are expessed in early fetal liver cells (Wong et al. 1983; Whitelaw et al. 1990).

The human globin loci have been studied more thoroughly, facilitated by the use of transgenic mice. The β LCR has been shown to be required for the initial activation of the locus and provides erythroid-specific, high-level, copy-number-dependent, position-independent expression to linked genes (Grosveld et al. 1987). Studies with the α locus have shown that HS-40 is required for expression of the α-like genes (Bernet et al. 1995) but transgenic results have suggested that additional sequences are required for developmental position-independent expression (Higgs et al. 1990; Sharpe et al. 1992; Gourdon et al. 1994). A detailed model of the mechanism of developmental regulation of the β-like genes has been proposed based on the observations of gene competition for LCR function (Enver et al. 1990; Hanscombe et al. 1991; Peterson and Stamatoyannopoulos 1993; Wijgerde et al. 1995, 1996; Dillon et al. 1997) and single gene

[1]Corresponding author.
E-MAIL fraser@ch1.fgg.eur.nl; FAX 31-10-436-0225.

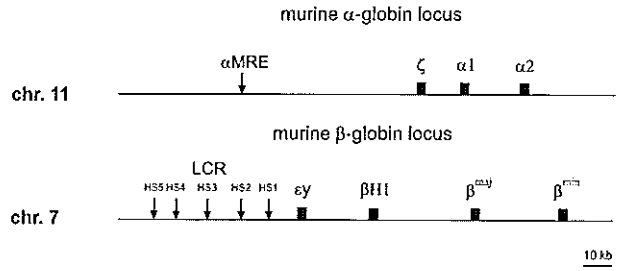


Figure 1. Schematic diagram of the murine *α-globin* and *β-globin* loci. Genes are represented by solid boxes and vertical arrows represent DNase I hypersensitive sites of the αMRE and β-globin LCR.

activation by the LCR (Wijgerde et al. 1995; Gribnau et al. 1998). These data argue strongly in favor of a looping model in which the LCR forms direct, mutually exclusive chromatin interactions with gene-local regulatory elements to activate transcription of a single gene. Multiple genes are coexpressed in *cis* through alternating transcriptional periods of single genes (Wijgerde et al. 1995; Gribnau et al. 1998) suggesting that LCR-gene interactions are dynamic but also semistable, persisting on the order of several minutes (Wijgerde et al. 1995; Dillon et al. 1997).

It has therefore been suggested that two parameters determine the transcriptional output and hence expression level of a given gene during development. The first is the frequency with which the LCR productively contacts a particular gene and the second is the stability of that interaction. The frequency of LCR–gene contact has been proposed to be dependent on distance from the LCR (Dillon et al. 1997). The relative distance between two competing genes and the LCR has been shown to be important in controlling both the level and timing of expression (Enver et al. 1990; Hanscombe et al. 1991; Peterson and Stamatoyannopoulos 1993). Dillon et al. (1997) measured the effects of distance on the frequency of LCR–gene interactions by comparing genes of equal stability at varying positions in the locus in combination with primary transcript in situ hybridization. Insertion of a β gene into more LCR-proximal positions resulted in that gene being transcriptionally activated more often and at the expense of the equivalent downstream gene in relation to the difference in distance.

The stability of the LCR–gene interaction has been proposed to be determined by the transcription factor environment. Targeted disruption of the erythroid Kruppel-like factor (EKLF) has shown that it is required for transcription of only the adult-type *β-globin* genes (Nuez et al. 1995; Perkins et al. 1995). EKLF binds selectively with high affinity to the CCACCC element present in the promoters of the mouse and human adult-type *β-globin* genes (Donze et al. 1995). Studies with compound human globin locus transgenic/EKLF knockout mice have shown that reductions in EKLF expression in heterozygous and homozygous knockout mice lead to decreased expression of *β-globin* and reciprocally increased expression of *γ-globin* mRNA (Wijgerde et al. 1996; Perkins et al. 1996). We have shown that these changes are caused by reductions in the number of transcriptionally active β genes in the fetal liver population with reciprocally increased numbers of active γ genes (Wijgerde et al. 1996). These studies have been interpreted to suggest that reduced EKLF levels lead to a decrease in the stability of the LCR–β gene complex. Reduction in the amount of time that the LCR complexes with the β gene allows more frequent interaction with the γ genes.

Here we present detailed characterization of the developmental expression and transcriptional regulation of the murine α- and *β-globin* loci at the single-cell level. The results of in situ hybridization and novel kinetic analyses suggest that transcriptional regulation of the mouse β genes is mediated by a similar dynamic chromatin interaction mechanism as has been proposed for the human β locus. However, unlike the human locus transcription of the adult β-like genes occurs in embryonic cells and hence is only suppressed partially and not silenced through competition for the LCR. Similar analyses of the *α-globin* locus indicate that a coactivational mechanism of multiple genes in *cis* cannot be ruled out.

Results

Developmental expression

RNA samples were collected from embryonic and fetal erythroid tissues at different stages of development and subjected to S1 nuclease protection assay to determine the expression pattern for the murine α- and *β-globin* genes (Fig. 2). Previous studies have suggested that small amounts of the adult *β-like* globin genes are expressed in embryonic cells (Brotherton et al. 1979; Chui et al. 1979; Wawrzyniak and Popp 1987; Whitelaw et al. 1990). Quantitative PhosphorImager analysis of S1 assays (Fig. 2A) shows that in 10.5-day embryonic blood the level of *εy* is 60% and *βH1* 34% of total *β-like* globin (Table 1). *βmaj–βmin* (the β S1 probe does not distinguish between *βmaj* and *βmin*) is detected at levels of ~6%. Expression of the α genes at 10.5 days is comprised of 46% ζ and 54% α mRNA (the α S1 probe does not distinguish between *α1* and *α2*) (Fig. 2B; Table 1). This is in agreement with previously published results (Whitelaw et al. 1990).

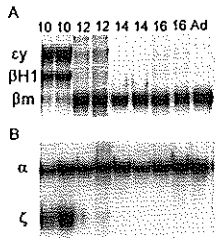


Figure 2. S1 nuclease protection assays. Total RNA from the indicated developmental timepoints (10.5-day whole embryo and 12.5-, 14.5-, 16.5-day fetal liver and adult blood) was subjected to S1 nuclease protection assay as described in Materials and Methods with probes of equal specific activity. (*A*) β locus probes *εy*, *βH1*, and *βm* (*βmaj* and *βmin*). (*B*) α locus probes ζ and α (*α1* and *α2*). The positions of protected fragments are indicated on the *left*.

At day 11.5 of gestation the main site of erythropoiesis changes from the embryonic yolk sac to the fetal liver, which is the site of definitive erythropoiesis. Analysis of fetal liver RNA at 12.5 days shows that 91% of the β-like mRNA is *βmaj–βmin*, and *εy* and *βH1* are still detectable at 7% and 2%, respectively. The appearance of small amounts of *βH1* and especially *εy* mRNA in the early fetal liver as late as 14.5 days has been suggested to indicate that expression of these genes is not restricted to yolk sac-derived cells (Whitelaw et al. 1990). However, this residual mRNA may also be caused by contamination of fetal livers with circulating embryonic cells. A definitive conclusion is only possible through individual cell analyses with cell-type markers and morphological analysis (see below). The levels of *α-globin* mRNA in 12.5-day fetal liver cells also shows a switch from roughly equal ζ and α expression in the embryonic cells to 99% α and only 1% ζ mRNA. ζ mRNA is undetectable by 14.5 days, suggesting but not proving that ζ transcription is restricted to embryonic cells.

Primary transcript in situ hybridization

We have shown previously that coexpression of the human γ- and *β-globin* genes in transgenic mouse fetal erythroid cells containing a single, complete human *β-globin* locus (Strouboulis et al. 1992) is achieved through alternating transcriptional periods of individual genes (Wijgerde et al. 1995). Primary transcript in situ hybridization with gene-specific intron probes showed that the vast majority of loci (88%) had only single gene signals. Approximately 12% of loci displayed two gene signals in *cis*, which we proposed represented a recent switch in gene activation. To ensure significance in this type of analysis the hybridization efficiency must be extremely high to guarantee the detection of nearly all transcriptionally active genes.

We determined the hybridization efficiency of the mouse ζ, α, *εy*, and *βH1* globin intron probes by hybridizing them to 10.5-day primitive erythroid cells (Fig. 3A,B; and Table 2). The ζ probes detect primary transcript signals at 96% of the loci. The α-globin intron probes that recognize the primary transcripts of both highly homologous *α1* and *α2-globin* genes demonstrate detectable signals at 89% of the loci (Table 2). Double-label experiments with both the ζ- and *α-globin* probes shows that a signal (α, ζ, or both) is detected at >95% of the loci. If all three *α-globin* genes are constitutively transcribed in primitive cells then our hybridization efficiencies are very high, at least >90%. It is possible that some of the genes are off at certain times; in this case we would conclude that our efficiency is even higher.

We performed similar quantitations for the *εy* and *βH1* globin primary transcript probes (Fig. 4B, Table 2). Approximately 91% of the loci have an *εy* signal and 63% have a *βH1* signal at 10.5 days. Calculation of the relative percent of *εy* versus *βH1* primary transcript signals (59% and 40%, respectively) yields comparable percentages to the relative amount of mRNA as determined by S1 analysis (60% and 34% of total β-like globin, respectively; Table 1). This close correlation between the relative percentages of transcriptionally active genes and mRNA expression suggests that the number of fully active genes in the population determines the level of mRNA expression. This was also found to be the case in the human β locus in which the results indicated that a gene is either fully on or off (Wijgerde et al. 1995, 1996; Dillon et al. 1997).

Globin gene transcription in primitive erythroid cells

Gene competition in the mouse *β-globin* locus has been suggested by a naturally occurring deletion of the *βmaj* gene and its gene-local regulatory elements (Skow et al. 1983) which leads to increased expression of the *βmin* gene in homozygotes (Curcio et al. 1986). In contrast, targeted disruption of the *βmaj* gene via insertion of a selectable marker leads to perinatal lethality with no increase in *βmin* expression (Shehee et al. 1993;). In the α locus, marker gene insertion into the ζ gene leads to reductions in the level of α gene expression in the fetal and adult erythroid cells suggesting that the marker gene may be competing for activation with the α genes (Leder et al. 1997).

To investigate the mechanism of coexpression in the

Table 1. *Quantitative S1 nuclease protection analysis*

	10.5 blood	12.5 fetal liver	14.5 fetal liver	16.5 fetal liver
εy	60	7	<1	0
βH1	34	2	<1	0
βmaj–βmin	6	91	99	100
ζ	46	1	0	0
α1–α2	54	99	100	100

PhosphorImager quantitation of S1 nuclease protection assays shown in Fig. 2. The numbers represent the percentage of total *β-globin* or *α-globin* mRNA expression for the individual β and α genes.

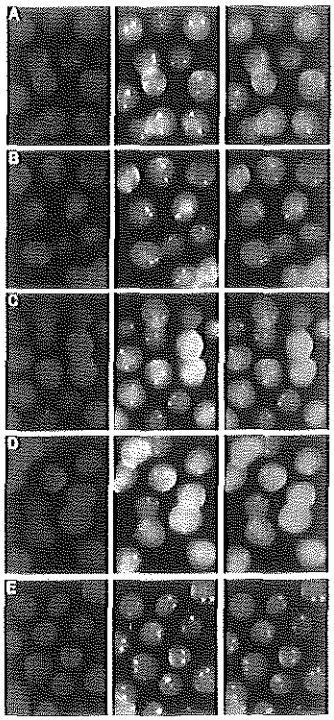


Figure 3. Primary transcript in situ hybridization on embryonic and fetal erythroid tissues. Double-label in situ hybridizations were performed with the indicated probe sets on 10.5-day blood (*A–D*) and 13.5-day fetal liver cells (*E*) as described in Material and Methods. Three separate images are shown for each hybridization; (*left*) red signal; (*right*) green signal; (*middle*) overlay of red and green. (*A*) ζ in red and α in green. (*B*) *βH1* in red and ϵy in green. (*C*) *βmaj* in red and ϵy in green. (*D*) *βmin* in red and ϵy in green. (*E*) *βmin* in red and *βmaj* in green. The photos are composites of CCD images that represent a single focal plane and therefore may not show both loci from an individual cell in focus.

mouse α- and *β-globin* loci and characterize the pattern of transcriptional regulation we performed primary transcript in situ hybridization, on 9.5-, 10.5-, and 11.5-day peripheral blood. Oligonucleotide probes specific for the introns of ϵy, *βH1*, *βmaj*, *βmin*, ζ, and α were used in double-label experiments. Primary transcript in situ hybridization shows that embryonic blood cells from days 9.5–11.5 transcribe all four β-like globin genes to varying degrees (Fig. 4A–C; Table 2). Each of the loci in a single cell can have a different primary transcript signal or combination of signals (Table 2) demonstrating that, like the human β genes, each locus responds independently to the same *trans*-acting factor environment. Almost all loci (91%; Table 2) have signals for the ϵy gene, which remains fairly constant in embryonic blood cells throughout the period from 9.5 to 11.5 days (Fig. 4A). The ϵy gene is closest to the LCR and is the most highly expressed globin gene during embryonic erythropoiesis (Fig. 2A). The ϵy and *βH1* foci appear frequently in *cis* (i.e., double signals on one locus) decreasing from 85% at 9.5 days to 45% at 11.5 days of gestation (Figs. 3B; 4A) as a result of the decrease in *βH1* gene transcription from 95% at 9.5 days to 53% at 11.5 days. Thus, fewer *βH1* genes are being transcribed as development proceeds.

Adult β gene transcription is observed in a small percentage of the *βmaj* genes in 9.5-day embryonic blood, increasing to 25% in 11.5-day blood (Figs. 3C; 4B). Three percent of the *βmin* genes had primary transcript signals at 10.5 days (Fig. 3D) and increased to ~10% in 11.5-day blood. The *βmaj* and *βmin* foci nearly always appear in *cis* with an ϵy signal in the early primitive cells, but by 11.5 days (blood) ~20% of the *βmaj* signals are single signals (Fig. 4B). These results of β gene transcription in the mouse globin locus are markedly different from those of the human transgene locus. The early transcription of the adult mouse β genes (*βmaj* and *βmin*) is in contrast to the human locus in which transcription of the adult *β-globin* gene is not detected in embryonic cells but is restricted to fetal-derived erythroid cells. These results show that the entire murine *β-globin* locus is activatable in embryonic erythroid cells.

In situ transcription analysis of the ζ and α genes in embryonic cells demonstrates that both genes are transcriptionally active as expected but also reveals a high percentage of double signals in *cis* starting at 90% in 9.5-day red cells and decreasing to ~70% at 11.5-days (Fig. 4C). The percentage of α genes with a signal remains fairly constant ~90% throughout the embryonic period but the percentage of ζ gene signals decreases from 99% of all ζ genes at 9.5 days to 80% at 11.5 days (blood), resulting in fewer double signals and more single α gene signals. As in the β locus the pattern of transcriptional activity of the genes in the α locus does not remain static from 9.5 to 11.5 days.

ϵy, βH1, *and* ζ *gene transcription is restricted to primitive erythroid cells*

The same probes were used to analyze gene transcription in fetal liver erythroid cells at 11.5 and 12.5 days of gestation (Fig. 4A–C). Livers were isolated from fetal mice at the time points indicated and distrupted gently and prepared for in situ hybridization as described in Materials and Methods. In situ analysis of 11.5- and 12.5-day fetal liver cells shows that the *βmaj* and *βmin* genes are now the most highly expressed genes, which is in agreement with the S1 analysis (Fig. 2A). At 11.5 days there is a small percentage of erythroid cells that continue to transcribe the ϵy, *βH1*, and ζ genes in the fetal liver preparations (Fig. 4A–C). It varies from one preparation to the next (average 15%) but by 12.5 days it is reduced to near zero. In situ analysis in 12.5-day peripheral blood in which >75% of the cells are nucleated primitive eryth-

Table 2. *Quantitation of transcriptional cell types and percentages of loci with transcription signals for the individual globin genes*

Transcriptional cell types		Cells	Loci with single, double, or not detected signals (%)				Transcriptional cell types		Cells	Loci with single, double, or not detected signals (%)				Transcriptional cell types		Cells	% loci with single, double, or not detected signals			
ϵy	$\beta H1$ ●	(%)	ϵy ○	$\beta H1$ ●	$\epsilon y/\beta H1$ ●	N.D.	●ζ	α○	(%)	ζ ●	α ○	ζ/α ●	N.D.	○β_{mul}	β_{min}●	(%)	β_{mul} ○	β_{min} ●	β_{mul}/β_{min} ●	N.D.
○		1	0	0.5	0	0.5	●		0.4	0.2	0	0	0.2	●		1	0	0.5	0	0.5
	●	<1	0	1	0	0	●	●	3	3	0	0	0	●	●	<1	0	1	0	0
●	●*	4	0	2	2	0	●	●*	13	6.5	0	6.5	0	●	●*	3	0	1.5	1.5	0
●	●	37	0	0	37	0	●	●	76	0	0	76	0	●	●	69	0	0	69	0
●	○*	31	15.5	0	15.5	0	●	○*	4	0	2	2	0	●	○*	14	7	0	7	0
●	○*	3	1.5	0	1.5	0	●	○*	2	1	1	0	0	●	○*	<1	0.5	0.5	0	0
●		6	0	0	3	3	●		0.45	0	0	0.2	0.2	●		8	0	0	4	4
○	○	14	14	0	0	0	○	○	1	0	1	0	0	○	○	2	2	0	0	0
○		3	1.5	0	0	1.5	○		0.15	0	0.1	0	0.1	○		2	1	0	0	1
		Total	31.5	3.5	59	5			Total	11	4	85	0.5			Total	10.5	3	81.5	5.5
	total % ϵy signals		31.5 +	0 +	59 +	0 – 91		total % α signals		0 +	4 +	85 +	0 – 89	total % β_{mul} signals			10.5 +	0 +	81.5 +	0 – 92
	total % βH1 signals		0 +	3.5 +	59 +	0 – 63		total % ζ signals		11 +	0 +	85 +	0 – 96	total % β_{min} signals			0 +	3 +	81.5 +	0 – 85

Double label in situ hybridizations as shown in Fig. 3 were counted and the percentages of the nine different transcriptional cell types are shown. (○, ●) The different primary transcript signals for the individual globin genes as indicated at the top of each column. The signals present on each of the homologous chromosomes are shown. (Overlapping ○, and ●). Loci with two different primary transcripts signals in *cis*. The percentages of individual loci with single, double or not detectable (N.D.) in situ signals contained within these cells are shown in the four columns on the right. The totals of these columns (single, double, or not detected) are shown directly beneath them. The appropriate values were summed up to arrive at the total percent of the specific gene signals (bottom two lines of each table). Results were taken from 10.5-day embryonic blood cells for ϵy, $\beta H1$, ζ and α. β_{mul} and β_{min} results are from 13.5-day fetal liver cells. Shown are the percentages of loci with transcription signals for the indicated genes.

*Cells with different transcription signals or combinations of signals on each locus.

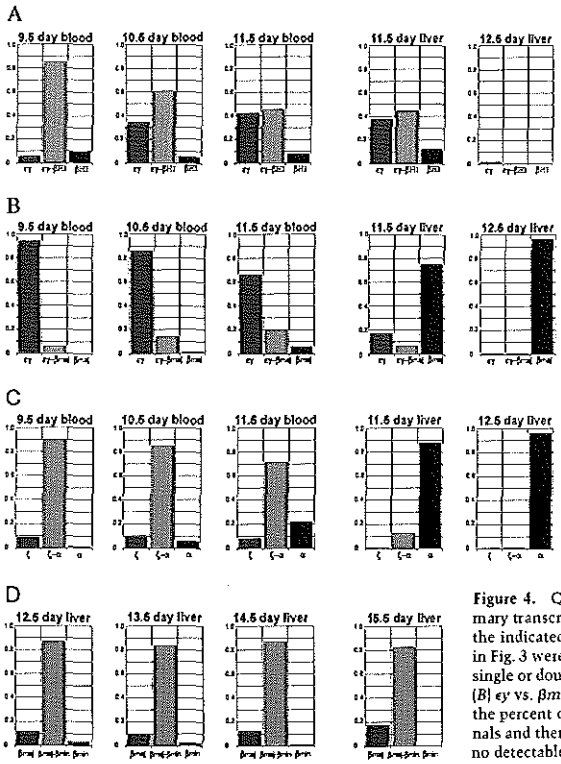


Figure 4. Quantitation of loci with single or double in situ primary transcript signals during embryonic and fetal development for the indicated genes. Double-label in situ hybridizations as shown in Fig. 3 were quantitated and plotted as the percentage of loci with single or double signals at the timepoints indicated. (*A*) ϵy vs. *βH1*. (*B*) ϵy vs. *βmaj*. (*C*) ζ vs. α. (*D*) *βmaj* vs. *βmin*. The values represent the percent of individual loci with detectable single or double signals and therefore do not include the small percentage of loci with no detectable signals (see Table 2).

rocytes shows that transcription of the globin genes is switched off dramatically (not shown). The near complete disappearance of ϵy, *βH1*, and ζ gene signals in the 12.5-day fetal liver preparations (Fig. 4A–C) suggested that those signals in the 11.5-day fetal liver preparations arose from contaminating embryonic erythrocytes. In addition the cells that display foci for the ϵy, *βH1*, and ζ genes in the 11.5-day fetal liver slides are distinct microscopically from the bulk of fetal liver cells as they display a high degree of autofluoresence, a characteristic of embryonic cells.

We confirmed that these cells were indeed embryonic and not fetal derived by three separate experiments. Hybridization of 11.5-day fetal liver slides with probes for ϵy, ζ, and *βmaj* show that a small percentage of autofluorescent erythroid cells (~15%) transcribe both the ϵy and *βH1* genes indicating that their expression is restricted to a subpopulation of cells in the fetal liver preparations (Fig. 5A–C). Transcription signals for mouse ζ and ϵy in cells from 11.5-day fetal-liver preparations from a homozygous transgenic line that contains a single copy of the human *β-globin* locus, were completely separate from the cells that had transcription signals for the fetal-restricted human β gene (Wijgerde et al. 1995) (not shown). Finally, in situ analysis of 11.5-day blood and fetal liver cells from an EKLF$^{-/-}$ mouse with ϵy and *βmaj* probes showed that the lack of *βmaj* transcription because of the EKLF knockout (Nuez et al. 1995; Perkins et al. 1995)

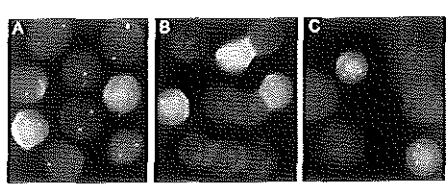


Figure 5. Primary transcript in situ analysis of 11.5-day fetal liver preparations. Cells were hybridized with (*A*) ζ in red and *βmaj* in green; (*B*) ϵy in green and *βmaj* in red; (*C*) ζ in red and ϵy in green. Note the relatively high autofluorescence in cells transcribing ϵy and/or ζ genes compared to the cells transcribing the *βmaj* gene.

did not lead to the appearance of ϵy transcription foci in the fetal liver cells. This indicates that the ϵy gene is silenced and not in competition with the *βmaj* gene in fetal cells (data not shown). Interestingly, *βmaj* transcription foci were not present in the EKLF$^{-/-}$ embryonic blood cells indicating that transcription of the *βmaj* gene is dependent on EKLF activity in primitive cells. We conclude from these experiments that the cells with mouse ζ and/or ϵy, and by inference those with *βH1* signals, are embryonically derived erythroid cells and that transcription of these genes is restricted to the primitive lineage and does not occur in true definitive cells.

Globin gene transcription in definitive erythroid cells

Presumably the mouse ζ, ϵy, and *βH1* genes are autonomously silenced during erythroid development as are the human ϵ (Raich et al. 1990) and ζ genes (Liebhaber et al. 1996). Definitive cells derived from the fetal liver express only the *α1* and *α2* genes from the α locus and the *βmaj* and *βmin* genes from the β locus. As mentioned previously the *α1* and *α2* genes are highly homologous even in their intervening sequences precluding the use of gene-specific intron probes to distinguish them.

The *βmaj* and *βmin* intron sequences are divergent and gene-specific probes were used to detect primary transcripts from these genes separately in double-label in situ hybridizations (Fig. 3E). Quantitation of both *βmaj* and *βmin* signals together shows that signals are detected at 95% of the loci in 13.5-day fetal liver indicating that the probe efficiency is very high (Table 2). Separate quantitation shows that 92% of the *βmaj* alleles and 85% of the *βmin* alleles have primary transcript signals (Table 2). Approximately 81% of the loci have *βmaj* and *βmin* signals in *cis* (double signals), 10% have *βmaj* alone and 3% have *βmin* alone (Table 2). A cell-by-cell analysis shows that ~17% of the cells transcribe a different combination of genes, whereas the *trans*-acting factor environment is the same (Table 2; indicated by *). The results from fetal liver cells from 12.5–15.5 days show that the percentage of *βmin* transcription foci is declining during development (Fig. 4D), which fits well with the previously reported changes in levels of β-maj and β-min proteins during fetal development (Whitney 1977; Alter and Goff 1980; Wawrzyniak and Popp 1987).

Coactivation or alternating single-gene activation

The relatively high percentage of double signals in the mouse α and β loci suggested that the mechanism of multigene expression in these loci may be different from that proposed for the human β locus. In the human β locus the LCR is thought to flip-flop between genes to alternately activate transcription. Primary transcript in situ hybridization analysis of γ and β transcription in the early fetal liver cells showed that 85% of the loci have single-gene signals (Wijgerde et al. 1995). The fact that nearly all cells had human γ and β mRNA in their cytoplasm indicated that alternation must occur and suggested that the small amount of double signals were caused by a recent switch from γ to β or vice versa. The results from the mouse α and β loci could be interpreted to indicate that transcription is coinitiated from multiple globin genes in *cis* in most cells. However, there are two indications from the data that contest this conclusion in the case of the β locus. First, a significant proportion of the cells have loci that are responding differently to the same *trans*-acting factor environment (38% at 10.5 days and 17% at 13.5 days, Table 2). Second, there is a significant proportion of loci with only single-gene signals (35% at 10.5 days and 14% at 13.5 days, Table 2). These results suggest the possibility that the individual genes in the mouse β locus may be alternating, albeit more often than in the human β locus (Wijgerde et al. 1995).

We therefore designed an experiment to provide further evidence of coinitiation or alternating transcription of multiple genes in the mouse globin loci making use of the inhibitor of transcriptional elongation DRB (5,6-dichloro-1-β-D-ribofuranosyl-benzimidazole). Previous studies have shown that DRB does not affect initiation of transcription (Fraser et al. 1978; Marshall and Price 1992) but prematurely aborts elongating transcripts 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). The effect of DRB treatment on globin gene transcription in mouse erythroleukemia cells has been reported previously (Tweeten and Molloy 1981). The results demonstrate that DRB causes premature termination without affecting initiation of transcription. This was tested in transgenic mice by briefly exposing fetal liver cells to DRB and analyzing initiation using promoter-proximal probes and elongation using probes to distal regions of the human γ- and β-*globin* primary transcripts via in situ hybridization (Gribnau et al. 1998). In situ signals with probes that hybridize to intron 1, located in the first 300 bases of the primary transcript are still visible after 15 min of DRB treatment, whereas probes that hybridize to intron 2, 600–1200 bases 3′ of the initiation site are no longer visible after 7 min of DRB treatment.

What makes DRB inhibition of elongation particularly useful is the fact that it is reversible, allowing elongation to proceed normally after removal of DRB. We tested this in day 10.5 embryonic blood cells by first treating cells with DRB for 15 min and then measured the reappearance of transcription elongation at intervals after washing out the DRB (Fig. 6). As a measure of recovery of the cells we recorded the percentage of cells in which both homologs had reappearing signals, using probes to intron 2 of the mouse ϵy and *βH1* primary transcripts. The results show that elongation signals reappear on both homologs in 67% of the cells, 2.5 min after washing out the DRB (Fig 6I; *trans* curve). This result illustrates two very important points. First, that elongation is restored very rapidly after removal of DRB and second, that the vast majority of cells that will recover already have elongation signals on both homologs. This shows that the elon-

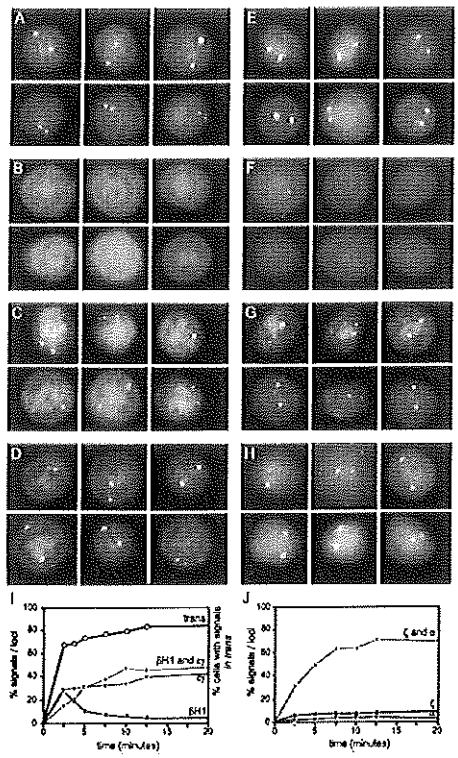


Figure 6. Kinetic analysis of reappearance of primary transcript signals after DRB inhibition of elongation. Blood cells from 10.5-day embryos were pulse treated with DRB as described in Materials and Methods and hybridized with mouse ϵy and *βH1* primary transcript probes (*A–D*), and mouse ζ and α globin primary transcript probes (*E–H*). (*A,E*) Prior to DRB treatment. (*B,F*) Cells treated with DRB for 15 min. (*C,G*) 5 min after washing out DRB. (*D,H*) 20 min after washing out DRB. The percent of single and double signals are presented from the indicated timepoints in *I* (ϵy and *βH1*) and *J* (ζ and α). Also shown is the percentage of cells with signals on both loci (*trans* curve; *I*) which is a measure of the recovery of cells after DRB treatment. All values represent the percentages of total loci or cells, respectively. The data represent the results of a single experiment was repeated four times in the case of β and three times in the case of α, yielding essentially the same curve shapes. Some variability is observed in the position of the curves along the *x*-axis, because of the inherent difficulties in obtaining short time points.

gation signals are not reappearing randomly throughout the available loci in the population of cells, but instead are contained in a subset of recovering cells in the early time points. The late time points show that the percentage of cells with *trans* elongation signals (signals on both homologs) are equal to untreated control values (82%), indicating that the cells are fully recovered and not adversely affected by the brief drug treatment.

We reasoned that if coinitiation in *cis* was occurring in loci that have double signals, then restoration of elongation after DRB treatment should result in the simultaneous reappearance of the same relative percentages of single- and double-gene signals. If alternation of transcription is the cause of the double signals, then single signals should reappear first and be followed by the reappearance of double signals after a lag that reflects the extra time required to produce a second signal in *cis*. Reappearing elongation signals were quantitated for ζ- and *α-globin* and ϵy and *βH1* globin genes at the indicated time points. The cells in the untreated panels (Fig. 6A,E) show the normal distribution of single versus double signals for the ϵy and *βH1* genes and the ζ and α genes, respectively. After 15 min of DRB treatment no primary transcript signals are detectable (Fig. 6B,F). Figure 6, C and G, shows an early time point after removal of the DRB block (5 min after wash). Note that the majority of signals in the case of the β genes are single-gene signals per locus, whereas for the α genes, most signals are double signals. A graphical analysis of the quantitation of loci with single- or double-gene signals as a percentage of all loci for the β and α genes are shown in Figure 6, I and J, respectively. In the case of the β genes, a higher than normal proportion of single signals are present in the early time points, whereas the double signals are underrepresented. As time progresses, the single signals are converted to double signals (most apparent from the *βH1* curve). The results for the α genes are strikingly different. The double and single signals reappear at approximately the same rate and with the same relative proportions as in control cells. The percentage of double signals does not significantly change in untreated control cells during the time course of the experiment (Fig. 7A). Comparison of the reappearance of double signals for the β and α genes by plotting the percentage of double signals/total signals, for DRB-treated cells normalized to control values shows that the double β signals are much slower to reappear when compared to the double α signals (Fig. 7B).

The fact that the relative amount of double α versus single α or ζ signals is the same as in the untreated cells, regardless of the number of cells that have recovered, suggests that a coinitiation mechanism could explain multigene expression in the mouse α locus. However, a transcriptional mechanism that involves very frequent alternations (i.e., alternations occurring every minute) would give the same result. The results support a mechanism of alternating transcription in the mouse *β-globin* locus that is similar to that proposed for the human *β-globin* locus (Wijgerde et al. 1995).

Discussion

We have used a combination of S1 nuclease protection assays on total RNA and primary transcript in situ hybridization in erythroid cells to analyze the gene expres-

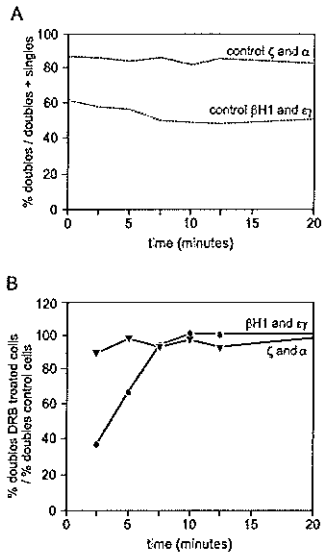


Figure 7. Percentage of double signals in control cells and reappearance of double signals in DRB-treated cells. (*A*) Double signals in untreated control cells are plotted as a percent of total signals for the α genes (ζ and α) and the β genes (εγ and *βH1*). (*B*) The reappearance of double signals in the DRB-treated cells are plotted as a percent of the control values in *A*.

sion and transcription patterns of the murine α and β globin genes. The results show that mRNAs from the embryonic εγ, *βH1*, and *ζ-globin* genes are expressed at high levels in primitive erythroid cells and are still detectable at low levels in the early fetal liver until 12.5–14.5 days of gestation. Primary transcript in situ hybridization indicates that the small amount of residual expression of these genes in early fetal liver preparations is caused by the presence of a subpopulation of cells that transcribe both the εγ and ζ genes. These cells display a high degree of autofluoresence and do not transcribe the fetal-restricted human β gene indicating that they are circulating embryonic cells and not true definitive cells derived from the fetal liver. We conclude from this data that transcription of the εγ, *βH1*, and *ζ-globin* genes is restricted to the primitive erythroid lineage.

In contrast, the *α1*, *α2*, and to a lesser extent the *βmaj* and *βmin* genes are expressed in the primitive lineage and are the only globin genes expressed in definitive cells. *α1* and *α2* mRNA makes up ~50% of the embryonic α globin at 10.5 days, whereas *βmaj* and *βmin* expression is ~6%. In the case of the β locus where all four genes are distinguishable, the results show that they are transcribed at frequencies that are inversely proportional to their distances from the LCR. These results are reminiscent of those of Dillon et al. (1997) in which a marked β gene was placed at different positions in the human globin locus and the effect of distance on transcription of the introduced gene and the other globin genes was measured. The results showed that altering the distance of a gene from the LCR had an effect on the frequency of transcription of that gene. An LCR-proximal gene was transcribed more often than a distal gene with reciprocal consequences for the remaining downstream genes in the locus. Those results suggested that the LCR activates gene transcription by forming direct chromatin interactions with gene-local regulatory elements and that the frequency of contact between two or more competing genes and the LCR is controlled in part by differences in distance.

The results highlight an important difference in the pattern of developmental regulation between the mouse and human β loci. In the mouse locus *βmaj* and *βmin* gene transcription is activated in the primitive lineage, whereas human β gene transcription is restricted to definitive erythroid cells in transgenics. The human β gene is thought to be silenced in primitive cells in part through competition for the LCR by the more LCR-proximal ε and γ genes. A competition mechanism could also be operating in the mouse *β-globin* locus, which is suggested by the DRB experiments, and could account for the observed inverse correlation between percentage of transcriptionally active genes and distance from the LCR in primitive cells. One might expect from these results that if the human β gene in its distal location in the locus is activatable in primitive erythroid cells then some transcription should be detected in situ. Clearly the human β gene is activatable in primitive cells when placed next to the LCR (Enver et al. 1990; Hanscombe et al. 1991) or in the position of the ε gene (Dillon et al. 1997) but not in its wild-type location (Wijgerde et al. 1995) or when placed just 5′ of the δ globin gene (Dillon et al. 1997). These results suggest the possibility that the distal part of the human locus that contains the δ and β genes is not accessible to LCR activation in primitive cells. This concept is supported by results from multicopy transgenic mice that contain head-to-tail tandemly integrated copies of a complete 70-kb human globin locus in which the β gene is juxtaposed on the 3′ side by an LCR (Milot et al. 1996). These mice do not express β mRNA in the primitive erythroid cells even though the β gene would be closer to the downstream LCR than the γ genes. It is important to note here that the LCR can activate the β gene in a 5-kb fragment when it is placed 5′ of HS5 (Zafarana et al. 1995) i.e., it works in both orientations. Taken together with the results presented in this paper that indicate that the *βmaj* and *βmin* genes are partially suppressed but not silenced by gene competition in primitive cells, it suggests that the human β gene is silenced in primitive cells through another mechanism that may involve epigenetic chromatin modification (Fig. 8). It has been proposed that the human β locus may be divided into distinct chromatin subdomains (see Collins and Weissman 1984 and references therein). This idea gains support from experiments with somatic cell hybrids between primitive transgenic ery-

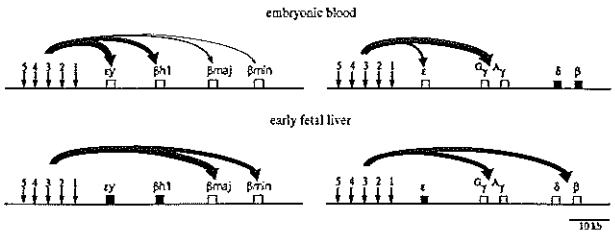


Figure 8. Models of developmental switching in the mouse and human β globin loci. Schematic diagrams of the mouse (*left*) and human (*right*) β globin loci are shown during embryonic (*top*) and early fetal (*bottom*) erythropoiesis. Vertical arrows denote DNase I hypersensitive sites of the LCRs. The curved arrows signify an interaction between the LCR and an individual globin gene. The relative thickness, of the arrows represents the percent of transcriptionally active genes in the population of erythroid cells. Open boxes represent activatable genes. Closed boxes represent genes that are silenced or in an inaccessible chromatin conformation (or both). Gray boxes (γ genes, fetal liver) reflect the possibility that the γ genes in some early fetal liver cells may be silenced (i.e., cells committed to β transcription; Wijgerde et al. 1996).

throid cells and MEL cells that show that γ gene expression is retained through several hybrid cell divisions before eventually switching to β expression (Papayannopoulou et al. 1986; Stanworth et al. 1995; see also Fraser et al. 1998). The above discussion is not intended to suggest that competition plays no role in transcriptional regulation of the *β-globin* genes. Competition is clearly operating in the human locus when multiple genes (or domains) are activatable, such as in primitive cells between ε and γ and in early definitive cells between γ and β.

The DRB experiments presented here suggest that alternating transcriptional activation of the mouse β genes is occurring, which may be indicative of gene competition for the LCR. Thus the results presented can be explained easily by the direct chromatin interaction mechanism between the LCR and an individual β gene that has been proposed for the human β locus. Alternative explanations would require the postulation of a number of other parameters. We propose a model in which the entire mouse β locus is accessible to LCR activation in primitive cells and that the individual genes are transcribed at frequencies that are inversely proportional to their distance from the LCR caused largely by gene competition (Fig. 8). In definitive cells the *εy* and *βH1* genes are no longer activatable because of silencing that may involve chromatin modification allowing the adult genes to be transcribed more frequently. Thus we propose that switching in the mouse locus involves shutting off the embryonic genes. In contrast, the switch to definitive erythropoiesis in the human transgene β locus involves 'opening up' of the adult δ–β domain through chromatin modification, making it accessible to activation by the LCR in competition with the γ genes.

The DRB results with the mouse α locus are more difficult to interpret but are clearly different from those of the β locus. The results of targeted disruption of the *ζ* and *α1* genes by insertion of a PGK–Neo cassette (Leder et al. 1997) are strongly indicative of gene competition for the αMRE. Insertion into the *ζ* gene had a more profound effect on *α1* and *α2* gene expression than did insertion into the *α1* gene, clearly indicating that relative position of the PGK–Neo cassette with respect to the αMRE is important. It has been firmly established that insertion of an activatable promoter (or gene) between a regulatory element and its normal target promoter leads to decreased activation of the downstream gene (Kim et al. 1992; Fiering et al. 1995; Hug et al. 1996; Pham et al. 1996; Dillon et al. 1997) indicating that gene order and/or relative distance from the regulatory element are key parameters in determining the competitive ability of a gene. We cannot conclude that the mouse α genes are alternately transcribed or coinitiated from our results, however the accumulated data suggest that they may be alternating frequently in a competitive mechanism for interaction with the αMRE. Obviously experiments that are able to show a direct interaction between the LCR or αMRE and an individual gene will allow firm conclusions regarding this type of mechanism.

Materials and methods

Preparation of RNA and S1 nuclease protection assay

RNA was prepared from 10.5-day embryos, 12.5- and 14.5-day fetal liver, and adult blood and subjected to S1 nuclease assay as described previously (Weaver and Weissman 1979). The *εy*, *βH1*, and adult β probes used were those described by Lindenbaum and Grosveld (1990). The mouse α and *ζ* probes were a 300-bp *Bam*HI fragment from plasmid GSE 1454 and a 300-bp *Apa*I–*Ava*II fragment subcloned from cosmid cML1 (Kielman et al. 1994), respectively. S1 nuclease-protected bands were quantitated with a PhosphorImager (Molecular Dynamics).

Primary transcript in situ hybridization

Primary transcript in situ hybridization to detect transcriptional activity of the mouse β-globin genes in 9.5-, 10.5-, and 11.5- blood; and 11.5-, 12.5-, and 13.5-day fetal liver cells was performed as described by Wijgerde et al. (1995). Embryonic blood cells were collected in PBS, fetal liver cells were disrupted in PBS (by pipetting up and down several times) and spotted on poly-L-lysine coated slides. Slides were placed in fixative (4% formaldehyde, 5% acetic acid in saline) for 20 min at room temperature then washed three times in PBS at room temperature and stored in 70% ethanol at −20°C. Slides were pretreated for hybridization by rinsing in Tris/saline (0.1 M Tris at pH 7 and 0.85% saline) and incubated in 0.01% pepsin, and 10 mM HCl at

37°C for 5 min. The slides were rinsed briefly in H_2O and fixed again for 5 min in 3.7% formaldehyde in PBS at room temperature. After a PBS wash slides were dehydrated in ethanol (70%, 90%, and 100%, respectively, 3 min each). Slides were hybridized in 25% formamide, 2×SSC, 1 mM EDTA, 5× Denhardt's, 50 mM NaH_2PO_4/$NaHPO_4$ (pH 7.0) and 200 ng of sheared salmon sperm DNA using 1–5 ng/μl of the appropriate oligonucleotide probes at 37°C overnight. After hybridization slides were washed in PBS for 30 min at 37°C and rinsed briefly in TST (Tris/saline with 0.05% TWEEN 20). The slides were then incubated in Tris/saline containing 1% blocking reagent (TSB) (Boehringer) for 30 min at room temperature. The slides were incubated subsequently with TSB containing the appropriate primary antibody (or Avidin D Texas red for biotinylated oligos) for 30 min at room temperature and washed two times for 5 min in TST. The previous steps were repeated with fluorescently labeled secondary (biotinylated goat anti-avidin D for biotinylated oligos) and tertiary antibodies. After the final washing step the slides were dehydrated (70%, 90%, and 100% ethanol), air-dried, and mounted with a 1:1 mixture of 1% DABCO, 0.4 μM DAPI (Sigma), 90% glycerol, 0.02% sodium azide, 10 mM Tris (pH 8.0) with Vectashield (Vector labs).

Quantitation of primary transcript signals was done by counting at least 400 cells for each timepoint using an epiflourescence microscope (Leitz). The figures presented were created with a CCD camera (Hamamatzu).

Oligonucleotide probes

Antisense oligonucleotide probes that recognize the intronic sequences of the respective globin gene primary transcripts were labeled with digoxygenin, dinitrophenol, or biotin haptens at both ends and in the middle as indicated. *εy* 1–4 (digoxygenin): CTCAGAATTCTTGATTTCCCTAGCTCTTTGTAC-CTAAAAAACAATTCTTCAGCCATTCACTGTCACCCTT-ACTGGGACCAATTAATTAACTTTGACAGCACTCTCTT-CATATTACTCTCCATATAAATCCATGATAAATTTTATC-ACGGTTGTCTTGCAAGACTTTTCTTCAACATCAATAA-ATAGGACCGCGCAAAA. *βH1* 1–4 (biotin): CAAAACCCT-ATAGAAACCCTGGAAATTTCTGCCATGCATAAGGATA-ATTTTGGACCCATGGACTCTAACATCTGTCAAGGCAT-TGCCAATCACAGTCTCAAAATGCTGGGCGCTCACTCA-AATCTGCACCCAAATCATTGTTGCCCACAAGCATAGA-TGTATTAATTTATAAAAACATACTCCTTTTTAAAAAAG-ATCCA. *βmaj* 2–4 (dinitrophenol): GAACTCTTGTCAACAC-TCCACACACAGTCATGGAGACTGCTCCCTAGAATATG-GGAAGTAAATAACCAGAGCTTAATTAATTTAGTAAAA-TGCAACTGGAGACAAATTATTATAAGAATCCTATGTC-AAACAGAATTTATATGTAAAATA. *βmin* 1–4 (biotin): TAT-GAAGTAGAGCAACAATACAAGATGCTGAAGGCCGAT-TTCAAATGGAAACTGTGGAAAGGATCAGAGAATCATT-TATCTTTTTGTCCTCAGAGTAAGAAAAAAAGAATTAT-TCTATGACACACAAAATTTAGCCACAAAATATACTCT-GGTAAAATGGCAGCTGGGTTCTACTGGTCAATTTTGA-TAAGAATTATTCT. *ζ1–5* (biotin): CCTTCTCAGTGGCTT-CTCCTCACTAACTGCTCTTTGTCACTTCTGTTCATGG-AAGACTCTGGTGAGCTCTGTGAATGCCAGCCCACCTC-CTTTAGTAACAACCCCAAGAGTGATGTTACTATTTGC-TGTTGCACAAGGGTCTACAAAGGGGATTTGATGCCTC-CAGCCTCCAATGGCACCCATGCCTGCGCTCG. *α1–3*(digoxygenin): TCACAGAAAAGCATAGTTAGAAGCTGCCCAC-TGAGCGAGTGCCAGGTCCATTAGCCCTTCCTAGGGGT-CCCAGATGCCGCCTGCCAGGTCCCTGCTCCCCTTCCTG-GGACCACTATGTTCCCTGCCTTGGGCACGAGGACCCT.

Reversible inhibition of elongation with DRB

Blood was collected from 10.5 day embryos in PBS and diluted to ~10^6 cells/ml. Half of the cells were treated with 100 μM DRB (5,6 dichloro-1-β-D-ribofuranosyl-benzimidazole, Sigma) for 15 min at 37°C, the other half were used as untreated controls. The cells were then diluted with several volumes of ice-cold PBS, spun at 200*g*, and washed two more times with ice-cold PBS and resuspended in PBS at ~10^6 cells/ml at 25°C. Aliquots were removed at 0, 2.5, 5, 7.5, 10, and 20 min after washing, spotted onto poly-L-lysine slides (Sigma) and prepared for in situ hybridization as described.

Acknowledgments

We thank S. Philipsen and A. van den Wijngaard for critical reading of this manuscript. We are grateful to M. Wijgerde, T McMorrow, E. Milot, M. Bender, and D. Higgs for helpful discussions and L. Braam for animal care. We are indebted to M. Kielman for the generous gift of the mouse α globin cosmid. T.T. is supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (N.W.O.). J.G. is supported by the Jeantet Foundation. This work is supported by the N.W.O.

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

Alter, B.P. and S.C. Goff. 1980. A murine model for the switch from fetal to adult hemoglobin during ontogeny. *Blood* **56:** 1100–1105.

Bernet, A., S. Sabatier, D. Picketts, R. Ouazana, F. Morle, D. Higgs, and J. Godet. 1995. Targeted inactivation of the major positive regulatory element (HS-40) of the human α-globin gene locus. *Blood* **86:** 1202–1211.

Brotherton, T., D.H. Chui, J. Gauldie, and M. Patterson. 1979. Hemoglobin ontogeny during normal mouse fetal development. *Proc. Natl. Acad. Sci.* **76:** 2853–2857.

Chodosh, L.A., A. Fire, M. Samuels, and P.A. Sharp. 1989. DRB inhibits transcription elongation by RNA polymerase II *in vitro*. *J. Biol. Chem.* **264:** 2250–2257.

Chui, D.H., T. Brotherton, and B. Gauldie. 1979. Hemoglobin ontogeny in fetal mice. Adult hemoglobin in yolk sac derived erythrocytes. In *Cellular and molecular regulation of hemoglobin switching* (ed. G. Stamatoyannopoulos and A. Nienhuis), pp. 213–225. Alan R. Liss, New York, NY.

Collins F.S. and S.M. Weissman. 1984. The molecular genetics of human hemoglobin. *Prog. Nucleic Acids Res. Mol. Biol.* **31:** 315–461.

Curcio, M.J., P. Kantoff, M.P. Schafer, W.F. Anderson, and B. Safer. 1986 Compensatory increase in levels of β^{minor} globin in murine β thalassemia is under translational control. *J. Biol. Chem.* **261:** 16126–16132.

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

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

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

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

Fraser, N.W., P.B. Sehgal, and J. Darnell. 1978. DRB-induced premature termination of late adenovirus transcription. *Nature* 272: 590–593.

Fraser, P., J. Gribnau, and T. Trimborn. 1998. Mechanisms of developmental regulation in globin loci. *Curr. Opin. Hematol.* 5: 139–144.

Gourdon, G., J. Sharpe, D. Wells, W. Wood, and D. Higgs. 1994. Analysis of a 70 kb segment of DNA containing the human ζ and α-globin genes linked to their regulatory element (HS-40) in transgenic mice. *Nucleic Acids Res.* 22: 4139–4147.

Gourdon, G., J. Sharpe, D. Higgs, and W. Wood. 1995. The mouse α-globin locus regulatory element. *Blood* 86: 766–775.

Gribnau, J., E. deBoer, T. Trimborn, M. Wijgerde, E. Milot, F. Grosveld, and P. Fraser. 1998. Chromatin interaction mechanism of transcriptional control *in vivo*. *EMBO J.* 17: 6020–6027.

Grosveld, F., G. Blom van Assendelft, D. Greaves, and G. Kolias. 1987. Position independent high-level expression the human β-globin gene. *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.

Higgs, D., W. Wood, A. Jarman, J. Sharpe, J. Lida, I. Pretorius, and H. Ayyub. 1990. A major positive regulatory region located far upstream of the human α globin gene locus. *Genes & Dev.* 4: 4679–4689.

Hug, B., A. Moon, and T. Ley. 1992. Structure and function of the murine β-globin locus control region 5' HS3. *Nucleic Acids Res.* 20: 5771–5778.

Hug, B., R. Wesselschmidt, S. Fiering, M. Bender, E. Epner, M. Groudine, and T. Ley. 1996. Analysis of mice containing a targeted deletion of β-globin locus control region 5' hypersensitive site 3. *Mol. Cell. Biol.* 16: 2906–2912.

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

Kielman, M.F., R. Smits, and L.F. Bernini. 1994. Localization and characterization of the mouse α globin locus control region. *Genomics* 21: 431–433.

Kim C.G., E.M. Epner, W.C. Forrester, and M. Groudine. 1992. Inactivation of the human β globin gene by targeted insertion into the β globin locus control region. *Genes & Dev.* 6: 928–937.

Leder, A., A. Kuo, M. Shen, and P. Leder. 1992. *In situ* hybridization reveals coexpression of embryonic and adult β-globin genes in the earliest murine erythrocyte progenitors. *Development* 116: 1041–1049.

Leder, A., C. Daugherty, B. Whitney, and P. Leder. 1997. Mouse zeta- and alpha-globin genes: Embryonic survival, alpha-thalassemia, and genetic background effects. *Blood* 90: 1275–1282.

Liebhaber, S., Z. Wang, F. Cash, B. Monks, and E. Russel. 1996. Developmental silencing of the embryonic ζ-globin gene: Concerted action of the promoter and 3'-flanking region combined with stage-specific silencing by the transcribed segment. *Mol. Cell. Biol.* 16: 2637–2646.

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

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

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

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: 105–114.

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

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.

Papayannopoulou, T., M. Brice, and G. Stamatoyannopoulos. 1986. Analysis of human hemoglobin switching in MEL X human fetal erythroid cell hybrids. *Cell* 46: 469–476.

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

Perkins, A.C., K.M.L. Gaensler, and S.H. Orkin. 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.* 93: 12267–12271.

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.

Pham, C.T., D.M. MacIvor, B.A. Hug, J.W. Heusel, and T.J. Ley. 1996. Long range disruption of gene expression by a selectable marker cassette. *Proc. Natl. Acad. Sci.* 93: 13090–13095.

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.

Sharp, J., P. Can-Thomas, J. Lida, H. Ayyub, W. Wood, and D. Higgs. 1992. Analysis of the human α-globin upstream regulatory element (HS-40) in transgenic mice. *EMBO J.* 11: 4564–4570.

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

Skow, L., B. Burkhart, F. Johnson, R. Popp, S. Goldberg, W. Anderson, L. Barnett, and S. Lewis. 1983. A mouse model for β-thalassemia. *Cell* 34: 1043–1052.

Stanworth, S.J., N.A. Roberts, J.A. Sharpe, J.A. Sloane-Stanley, and W. Wood. 1995. Established epigenetic modifications determine the expression of developmentally regulated globin genes in somatic cell hybrids. *Mol. Cell. Biol.* 15: 3969–3978.

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.

Tweeten, K.A. and G.R. Molloy. 1981. Induction of premature termination of transcription of the mouse β globin gene by DRB. *Nucleic Acids Res.* 9: 3307–3319.

Wawrzyniak, C. and R. Popp. 1987. Expression of the two adult β-globin genes in mouse yolk sac and fetal liver erythrocytes. *Dev. Biol.* 119: 299–301.

Weaver, R.F. and C. Weissman. 1979. Mapping of RNA by a modification of th Berk-Sharp procedure: The 5' termini of

15s β globin mRNA precursor and mature 10s β globin mRNA have identical map co-ordinates. *Nucleic Acids Res.* **6:** 1175–1193.

Whitelaw, E., S. Tsai, P. Hogben, and S. Orkin. 1990. Regulated expression of globin chains and erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse. *Mol. Cell Biol.* **10:** 6596–6606.

Whitney, J.B. 1977. Differential control of the synthesis of two hemoglobin β chains in normal mice. *Cell* **12:** 863–871.

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

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

Wong, P.M.C., S. Chung, J.S. White, S.M. Reicheld, M. Patterson, B.J. Clarke, and D.H.K. Chui. 1983. Adult hemoglobins are synthesized in murine fetal hepatic erythropoietic cells. *Blood* **62:** 1280–1288.

Zafarana, G., S. Raguz, S. Pruzina, F. Grosveld and D. Meijer. 1995. The regulation of human β-globin gene expression: The analysis of hypersensitive site 5 (HS5) in the LCR. In *Molecular biology of hemoglobin switching. Proceedings of the Ninth Conference on Hemoglobin Switching* (ed. G. Stamatoyannopoulos), pp. 39–44. Intercept Ltd., Andover, Hants, UK.

Chapter 6.

Developmental activation of chromatin domains in the human β globin locus.

Parts of this chapter will be submitted.

Developmental Activation of Chromatin Domains in the Human β Globin Locus.

Abstract

We have investigated intergenic transcription and chromatin structure in the human β globin locus during erythroid development in transgenic mice. *In situ* hybridization studies demarcate intergenic regions of the locus which are transcribed in a developmentally and cell cycle specific manner. Developmental analysis of DNase I general sensitivity across the locus reveals sharply defined, differentially activated, chromatin sub-domains which precisely correspond to the areas delineated by intergenic transcription. The results show that the globin locus is divided into at least 3 distinct chromatin subdomains, two of which are differentially opened and closed during developmental progression. Deletion of the sequence element that segregates the adjacent embryonic and adult domains results in an activation failure of the β globin gene (Calzolari 1999). We show that this deletion results in the abrogation of intergenic transcription in the adult domain and suggest a link between intergenic transcription and maintenance and/or activation of chromatin structural domains.

Introduction

The human β globin locus control region (LCR) was functionally defined by its ability to confer tissue specific, high level expression upon linked genes in transgenic mice and erythroid cell lines, independent of the position of integration in the host genome (Grosveld et al., 1987). Transgenic mice with human globin gene sequences without a linked LCR fail to express these genes. Likewise patients with naturally occurring large deletions including the LCR but otherwise normal globin genes exhibit $(\varepsilon\gamma\delta\beta)^0$ thalassemia. The genes in such loci are transcriptionally inactive, DNase I insensitive, and late replicating. Deletion of individual hypersensitive site regions in the LCR in transgenic mice containing full locus constructs leads to loss of position independence (Milot et al., 1996; Li et al., 1998). In view of these properties, it was proposed that LCR's were required to open or activate the chromatin of the globin locus in red cells.

Recent results using chromosome transfer and recombinase mediated removal of the human LCR (HS:2-HS:5) in an erythroid cell environment indicate that it is not required to maintain a DNase I sensitive state (Reik et al., 1998). Similar experiments in which the mouse β globin LCR was deleted after chromosome transfer to human K562 erythroleukemia cells showed equivalent DNase I sensitivity of the mouse and endogenous human β globin loci (Epner et al., 1998). Further experiments in which the murine LCR was deleted prior to chromosome transfer from ES cells into K562 cells showed equivalent sensitivity in the regions of the adult human and mouse β globin genes. These results were interpreted to indicate that the LCR is not required for chromatin activation or maintenance of activation and that other elements are sufficient for these functions.

We have previously shown that the human and murine LCR's and intergenic regions are transcribed in a subset of erythroid cells *in vivo* (Ashe et al., 1997). Intergenic transcription in these regions is unidirectional and RNA polymerase II specific. Recently, an LTR retrotransposon has been identified about 2 kb upstream of HS:5 of the human LCR (Long et al., 1998). LCR transcripts initiate in the LTR and terminate in the region of the ε gene silencer located approximately 23 kb

downstream (Proudfoot personal comm.). Interestingly, the smallest of the LCR thalassemia deletions [the Hispanic (εγδβ)° thalassemia] removes ~30 kb including HS:2-5 of the LCR and ~20 kb of upstream sequences including the LTR. This region is capable of partial insulation of a globin gene from the LCR in transgenic mice (Zafarana et al., 1995) and represents the only differences in *cis* elements between the deletion used by Reik et al., (1998) which maintained chromatin structure in cell lines and that of the thalassemia patients which results in complete inactivation.

Results

β globin LCR and intergenic transcription during development

We mapped the extent of intergenic transcription *in situ* in transgenic lines containing the entire human β globin locus (Strouboulis et al., 1992). In transgenic embryonic red cells the human ε and γ genes are transcribed, whereas the δ and β genes are transcriptionally silent (Wijgerde et al., 1995). Intergenic transcripts which colocalize with ε and/or γ gene transcripts are detectable *in situ* at approximately 10-15% of globin loci in embryonic red cell nuclei (E 10.5) with probes c, d, and e but not with probes f, g and h (Figure 1). LCR transcripts are detectable with probes a and b in embryonic red cells and red cells from all other developmental stages (this paper and Proudfoot pers. comm.). The results presented here suggest that transcripts are present throughout the intergenic regions of the ε and γ genes and continued for at least another 10 kb downstream of the γ genes. The fact that they are not detectable with probes f, g and h in the region of the δ and β genes suggests that a barrier to, or terminator of intergenic transcription resides in the region between probes e and f in embryonic red cells, since probes 5' of this region detect transcripts but downstream probes do not.

In early fetal liver derived red cells (E 12.5) the ε gene is silenced by a proximal upstream silencer element (Raich et al., 1990), and γ gene transcription persists in competition with the δ and β genes (Wijgerde et al., 1995, 1996). The LCR is thought to alternately activate transcription of individual globin genes *in cis* to bring about expression of both the γ and β genes in single erythroid cells (Wijgerde et al., 1995; Dillon et al., 1997; Gribnau et al., 1998). *In situ* analysis with all probes (a-h; LCR and throughout the locus) in E12.5 red cells detects intergenic transcripts colocalizing with 10-15% of globin loci which are actively transcribing a globin gene, although transcription in the ε and γ regions are reduced compared to embryonic cells. Importantly, intergenic transcription signals are detectable with probes flanking the β gene in loci which are not at that moment transcribing the β gene (Ashe et al., 1997; and not shown). Since 95% of the loci are actively transcribing a γ or β gene (Wijgerde et al., 1995) the γ gene is being transcribed at that moment on those loci. Similarly, intergenic signals detected with probes in the δβ region can be found *in cis* with loci transcribing the γ genes. Double label hybridizations with intergenic probes in the εγ and δβ regions show that transcripts for both regions are detectable *in cis*, demonstrating that the entire locus is transcribed in individual early fetal red cells. These results show that intergenic transcription is not simply a consequence of adjacent genic transcription, but rather reflects the potential of the genes in a region to be transcriptionally activated.

In adult red cells the ε and γ genes are silenced and the β gene is expressed at high levels. Silencing of the γ genes is thought to occur via a complex mechanism including multiple silencing elements (see below). *In situ* analysis of

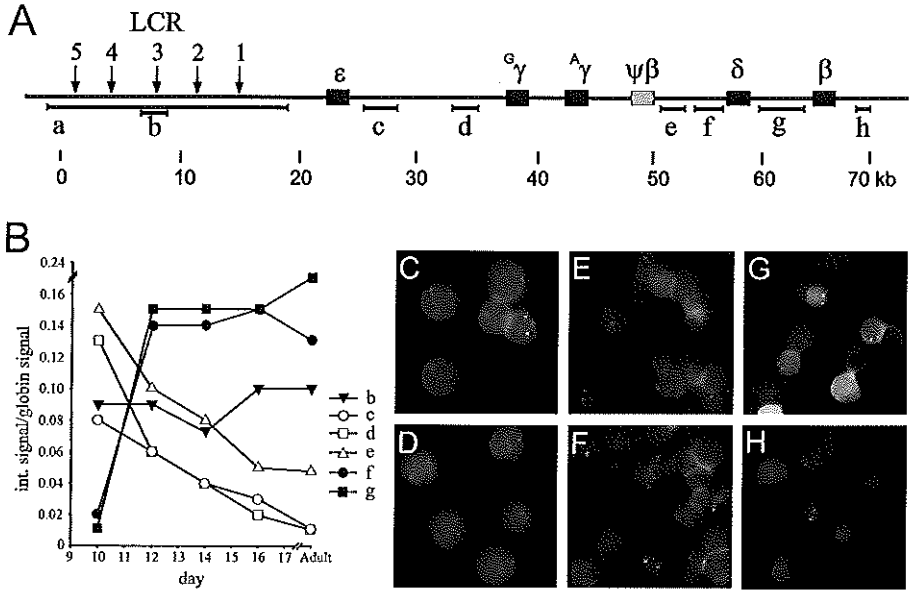


Figure 1. *Panel A shows a schematic diagram of the localization of the probes that have been used for the detection of intergenic transcripts by in situ hybridization using cells from different developmental stages. Intergenic signals using different probes were counted and plotted during development (B). Embryonic blood cells show positive signals for probe c (C) but fail to do so for probe g (D), green foci (FITC) represent intergenic signals and red foci (Texas-red) γ gene signals. A severely reduced amount of signals is detected with probe c in anemic spleen cells (E) whereas probe g now show positive signals (F), green signals represent intergenic signals and red signals β gene transcription. Panel G shows co-detection of probes c,d in red and f,g in green on 12.5 day fetal liver cells. Panel H, intergenic transcripts detected with probe c (green) in cord blood cells, red signals are γ and β gene signals.*

adult anemic spleen red cells reveals that intergenic transcripts are detected at high levels (10-25% of loci) only with probes a, b, f, g and h, in the region of the LCR and flanking the δ and β genes. Signals with probes c, d, and e (in the ε–γ region) have continued their downward trend and are now barely detectable. If one assumes that intergenic transcripts are merely the result of fortuitous transcription of open or active chromatin domains then the results presented here lead to several important conclusions. First of all the globin locus appears to be divided into at least 3 chromatin domains. The first domain, containing the LCR is in a developmentally stable open conformation in red cells. The εγ domain is open in embryonic and early fetal red cells and closes during late fetal and adult development. The δβ domain is closed during embryonic erythropoiesis, opens in fetal red cells and stays open during adult erythropoiesis. Importantly, in early fetal liver red cells, intergenic signals are detectable in both the εγ and δβ domains *in cis* indicating that the entire locus is open at this stage. This is the period of development when the γ and β genes compete for LCR activation (Wijgerde et al., 1995).

Chromatin structure

The data suggest that the region between probes e and f contains a significant functional element. It represents the 3' border of intergenic transcription in the εγ domain in embryonic cells and the 5' border of intergenic transcription in the δβ domain in adult cells. If intergenic transcription indeed delineates areas of active chromatin we would expect to see a transition from DNase sensitive to DNase insensitive chromatin on either side of this region which should be dependent on the developmental stage. We therefore performed general DNase sensitivity assays on isolated nuclei from E10.5 blood and adult anemic spleen from two transgenic lines containing the entire human β globin locus. Necessarily we avoided analyzing fragments that contain hypersensitive sites, and wherever possible compared fragments of equivalent sizes. The relative sensitivity of each fragment was determined by calculating the sensitivity factor (S) according to Wood and Felsenfeld, (1982) which corrects for size differences. In embryonic cells, restriction fragments in the εγ domain are equivalent in sensitivity to fragments in the LCR, and are 6 fold more sensitive to DNase than the inactive zfp37 gene (Mazarakis et al., 1996) as expected. However, fragments in the δβ domain, were 2-3 fold less sensitive than the εγ domain and LCR. The transition from the DNase sensitive εγ domain to the relatively insensitive δβ domain appears to lie in a fairly circumscribed region (compare fragments 3'ψβ and 5'δ in Fig. 2C). These results parallel the intergenic transcription results which showed a 3' limit to intergenic transcription in the same region in embryonic cells.

In adult red cells fragments in the δβ domain are equivalent in sensitivity to the LCR and are 2-3 fold more sensitive to fragments in the εγ domain. Here again the pivotal region lies between probes e and f (compare fragments 3'ψβ and 5'δa in Fig. 2C) indicating that this region contains a chromatin border. These results show a clear-cut correlation between chromatin structure and intergenic transcription, indicating that domains of active chromatin are delineated by intergenic transcription. Furthermore the results show that the globin locus is not a single open chromatin domain in red cells, rather it appears to be a series of at least three chromatin domains, two of which are developmentally activated and silenced. Perhaps most important is the fact that a region of greater than 30 kb of relatively closed chromatin (εγ domain) between the LCR and δβ domain in adult red cells does not affect LCR directed transcription of the β gene.

Analysis of intergenic transcription in transgenics with deleted 5'δβ domain border

Several point mutations and small deletions in the promoter regions of the γ genes have been identified which cause hereditary persistence of fetal hemoglobin (HPFH) which phenotypically manifests as pancellular elevated γ gene expression in adult red cells. Interestingly HPFH can also be caused by specific large deletions in the globin locus which remove the δ and β genes and several kb of 3' flanking sequences, juxtaposing distant enhancer elements with the globin locus. Many other very similar large deletions of the δ and β genes have been described which result in δβ thalassemia. Whether a deletion leads to HPFH or δβ thalassemia seems to depend on the precise location of both the 5' and 3' break points (see discussion). The minimal region of difference between the 5' breakpoints of an HPFH deletion and a thalassemia deletion is approximately 900 bp, located approximately 2.8 to 3.7 kb upstream of the δ gene (Collins and Weissman, 1984). Among the proposals to

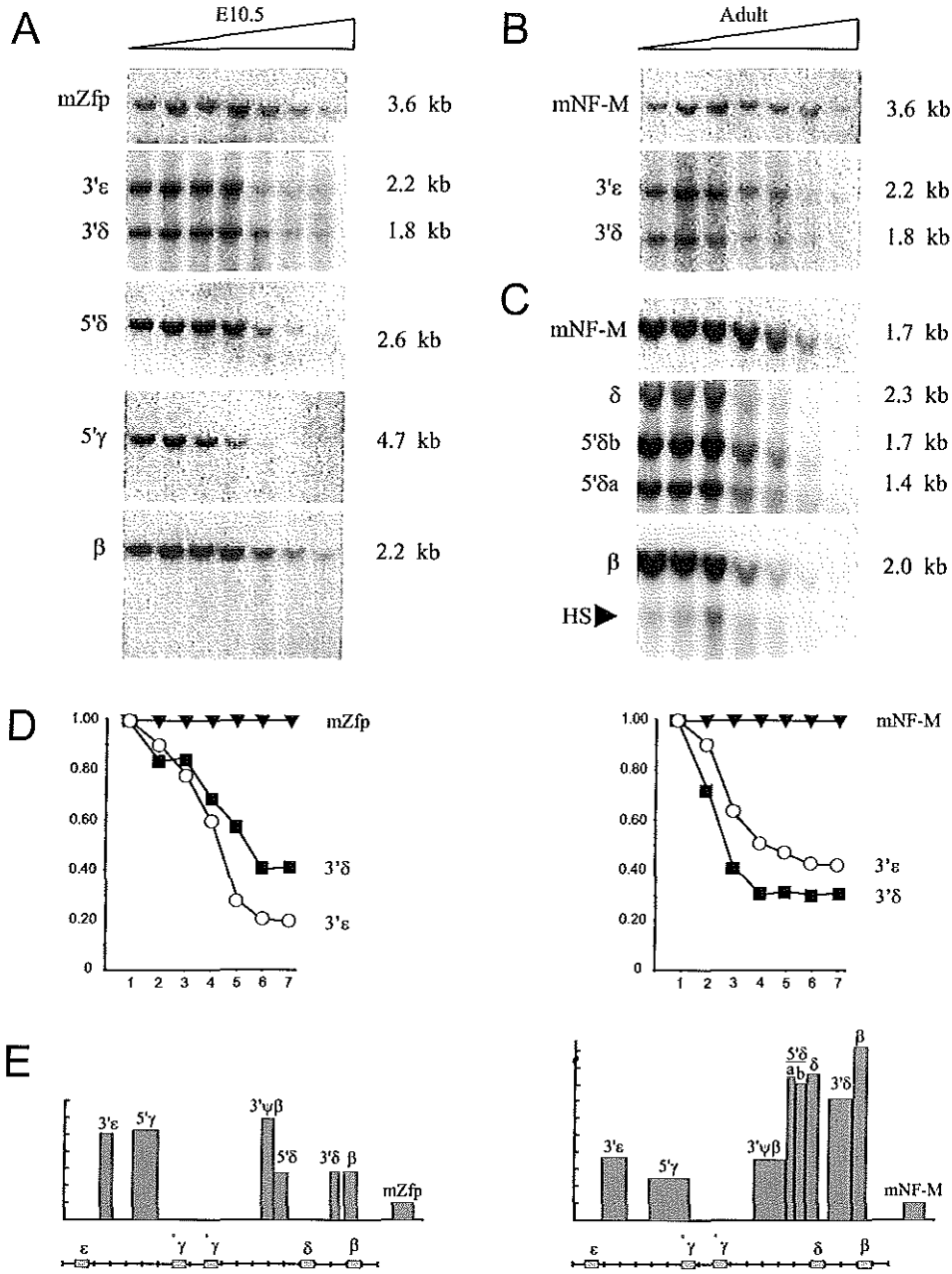


Figure 2. General DNase I sensitivity analysis during development. (A) Different restriction fragments (HindII-AlwNI) were tested for general DNase I sensitivity in 10.5 day embryonic blood cells and anemic spleen cells (B). (C) Sensitivity of different fragments (EcoRI-SphI) in the region in front of the β gene. (D) Quantitation of DNase I digestion levels of fragments within the εγ and the δβ domains with increasing DNase I concentration is plotted relative to the digestion level of a fragment in the insensitive NF-M and Zfp genes. (E) The relative sensitivity of the tested fragments corrected for size according to Wood and Felsenfeld (1982).

explain the phenotypic differences caused by the various deletions was the idea that the γ and β genes were located in separate chromosomal domains (Bernards and Flavell 1980). The chromatin border that we have identified appears to be located in the 5'δa fragment (Figure 2C) which completely overlaps the minimal region of difference.

Based on the patient data two transgenic lines were created containing a human β globin YAC with a 2.5 kb deletion (Δ2.5 YAC) which encompasses the minimal initiation region (Calzolari et al., 1999). Embryonic red cells from these mice express near normal levels of ε and γ globin mRNA. However, expression of the β gene in fetal and adult red cells is variegated with only 20% of the cells expressing the human transgene. In addition it was shown that the level of transcription per active gene is reduced to approximately 50% of normal. These results indicate that this region is essential for normal activation of the β globin gene in the context of the full locus.

In situ analysis of intergenic transcription in embryonic cells from the Δ2.5 YAC line 3 which contains a single copy of the mutated YAC shows normal levels of transcripts in the LCR and εγ domains with the exception of the region just upstream of the deletion (probe e) which is reduced from 15% of loci to 1%. In E14.5 fetal red cells intergenic transcription in the εγ domain decreases normally however we observe a near absence (≤1%) of intergenic transcription in the δβ domain suggesting that the chromatin is closed. DNase I analysis shows that the sensitivity of the δβ domain is reduced to that of the inactive zfp37 gene (approx. 1.2 fold). These results show that the deleted region is essential for initiation of intergenic transcription and chromatin activation of the δβ domain. Intergenic transcription in the LCR is also reduced but persists, as it does at low levels in both the εγ and δβ domains. However, these transcripts are present only in a fraction of the expressing pool of variegated erythroid cells. This suggests that in the absence of an active δβ domain (or β gene) the chromatin of the LCR also closes down, which is in agreement with the results of Reitman *et al.*, (1993) which showed that the LCR alone is not capable of maintaining an active chromatin structure.

Intergenic transcription occurs in specific phases of the cell cycle

We were interested in determining why only a subset of globin loci had intergenic transcription signals. We noticed that in homozygous transgenic mice, greater than 50% of the intergenic signals are in cells which display intergenic signals on both loci. Analysis of intergenic transcription in the endogenous mouse globin locus and the endogenous human locus in cord blood BFU-E cultures showed similar co-ordination of loci in the same cell (figure 1H). These results indicate that intergenic transcription is not a random phenomenon and suggested that it occurs in response to a specific intercellular environment.

In order to investigate the timing of intergenic transcription we performed *in situ* hybridization on FAC sorted G1, early S, late S and G2 cells from E13.5 fetal liver. The results demonstrate that in a random population of 100 cells 17% had intergenic foci and were in G1, 2% with foci were in early S, 1% were in late S and <1% were in G2, indicating that the bulk of intergenic transcription occurs in G1 cells (figure 3).

We considered the possibility that the small amount of intergenic foci detected in S and G2 phase cells was the consequence of contamination of G1 phase cells. We tested this by marking S phase cells in random populations with BrdU

incorporation and immunofluorescent detection combined with *in situ* hybridization. Approximately 6% of the BrdU labelled S phase cells (2% of total) had intergenic signals confirming that a small population of S phase cells has intergenic transcription, while the majority occurs in non-S phase cells.

To further investigate the timing of intergenic transcription within the population of S phase cells we combined immunofluorescence detection of proliferating cell nuclear antigen (PCNA) with *in situ* hybridization detection of

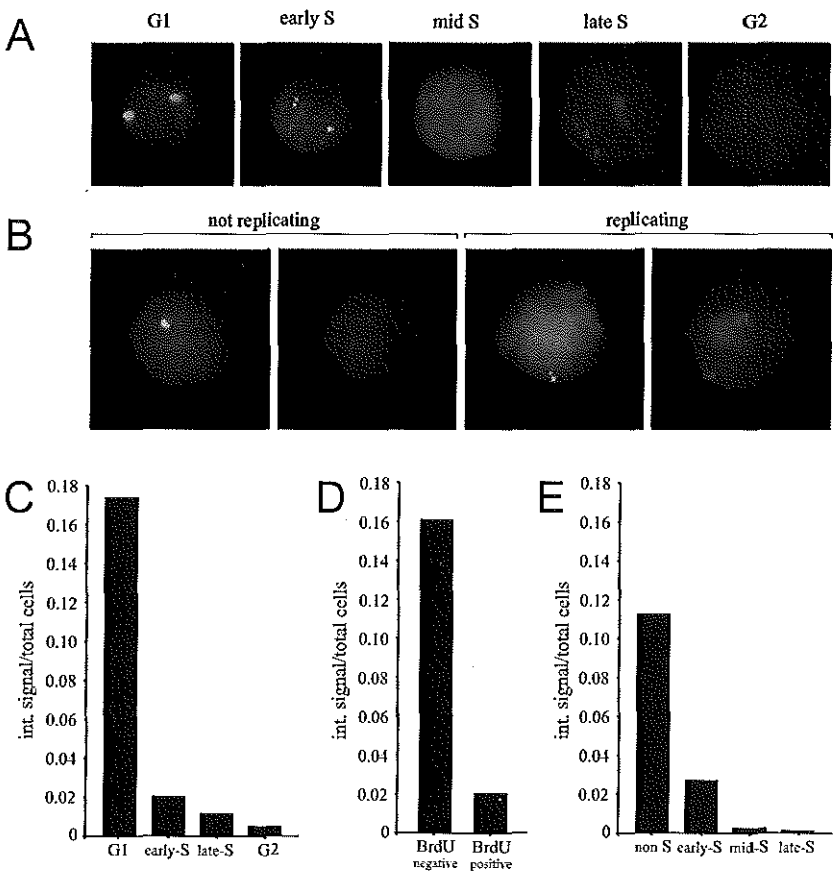


Figure 3. *Identification of intergenic transcription during the cell cycle. Panel A shows detection of intergenic transcription foci in green (FITC) using probes b,c,d,e,f and g in conjunction with detection of the cell cycle marker PCNA (Texas-red) in 13.5 day homozygous fetal liver cells. Panel B shows the detection of intergenic transcription foci in green using probes b,c,d,e,f and g in combination with BrdU staining (Texas-red) in 13.5 day heterozygous fetal liver cells. Quantitation of the amount of intergenic foci at different stages during the cell cycle as determined by cell sorting (C), BrdU staining (D) and PCNA staining (E). The bar graphs represent the distribution of intergenic signals over subsets of a random population of cells.*

intergenic transcripts. PCNA changes its subcellular localization during the cell cycle making it useful for marking cells at specific stages of the cell cycle (Humbert et al., 1992). It is particularly useful in discriminating among early, mid and late S phase cells. In early S phase PCNA staining appears as small foci in the interior of the nucleus. During S phase progression the number of PCNA foci first increases and later decreases. In late S phase the size of the clusters increases and are located predominantly in the nuclear periphery. The *in situ* hybridization data reveals that the intergenic transcription in S phase cells occurs only during early S. No intergenic transcription foci were detected in mid or late S phase cells. We conclude that intergenic transcription occurs only in early S and G1 cells. The small amount of cells with intergenic signals in late S and G2 in the sorted populations is most likely the result of contamination.

Discussion

We have shown that the human β globin locus is actually divided into at least three chromatin domains through the analysis of intergenic transcription and general DNase I sensitivity. The LCR domain is transcribed in red cells at all developmental stages and is constitutively sensitive to DNase I. The rest of the locus is divided into two developmentally specific chromatin domains. In embryonic red cells the εγ domain is transcribed and is 2-3 fold more sensitive to DNase I than the non-transcribed δβ domain. In early fetal red cells εγ domain transcription decreases as δβ domain transcription increases, however transcripts for both domains can be found *in cis* suggesting that the entire locus is open and active at this stage. During adult erythropoiesis transcripts in the εγ domain are severely reduced whereas the δβ domain is transcribed at a high level and is 2-3 fold more sensitive to DNase than the εγ domain. The transition or border between DNase I sensitive and insensitive domains in both embryonic and adult cells is located in an area between -2.6 and -4.8 kb upstream of the δ gene. This area includes the 900 bp region (-2.7 to -3.6) which is the minimal difference between a $(\delta\beta)^{\circ}$ thalassemia deletion and an HPFH deletion. This region was originally proposed to be a domain border based on analysis of the breakpoints in various deletions of the δ and β genes and their resultant phenotypes. Deletions of the genes and the putative border element displayed persistent, pancellular expression of the γ genes in adult red cells presumably due to juxtaposition of active downstream sequences (Anagnou et al., 1995). Deletions with downstream 5' breakpoints, the juxtaposed sequences had no effect on γ gene expression which were silenced normally. The border region is also contained within the 2.5 kb deleted region (-2.4 to -4.9) analysed by Calzolari et al., (1999). which they have shown to be essential for activation of the β gene.

The domain organisation of the globin locus has a number of important implications in understanding developmental control of transcription of the β genes. First, it appears that gene competition for the LCR is not a key factor for the absence of β gene expression in embryonic cells. Previously it was thought that competition from the ε and γ genes, which are closer to the LCR, suppressed β gene transcription in embryonic cells. Here we show a sharply defined domain of relative closed chromatin surrounding the δ and β genes which suggests that the lack of β gene transcription in embryonic cells, is controlled by chromatin inaccessibility. Our results suggest that only genes in areas of active chromatin can compete for LCR activation of transcription. It is known that EKLF, one of the principle factors controlling β gene transcription (Nuez et al., 1995; Perkins et al., 1995, 1996;

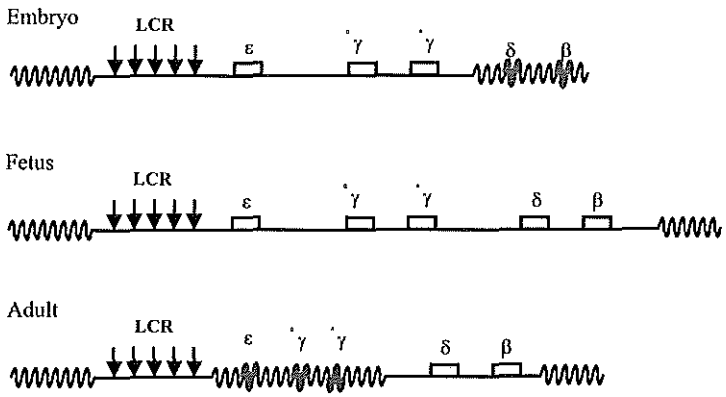


Figure 4. *A model for the regulation of the activation of the globin genes during development. The β globin locus is divided into three seperate domains. The LCR domain is in a constitutively open chromatin configuration during development. The εγ domain is open in embryonic and early fetal erythroid cells. The δβ domain opens in early fetal erythroid cells in cis with the εγ domain. In adult cells the εγ domain closes. Genes that reside in domains with an active chromatin configuration are able to compete for activation by the LCR.. The curled lines represent inactive chromatin whereas straight lines delineate active chromatin.*

Wijgerde et al., 1996) is readily available (Southwood et al., 1996) and active in embryonic cells (Tewari et al., 1998; Trimborn et al., 1999) however, it is not able to affect the human β gene unless it is taken out of the context of its domain (Tewari et al., 1998).

This is in contrast to the murine homologues of the human β gene; the mouse β^{maj} and β^{min} globin genes are expressed in embryonic cells in an EKLF dependent manner (Trimborn et al., 1999). All four mouse β globin genes are active in embryonic cells at levels that are inversely proportional to their distance from the LCR suggesting that the entire locus is activated and all of the genes are competing for the LCR. This also appears to occur in the human locus in early fetal cells when both the εγ and δβ domains are activated *in cis*. At this stage dynamic, mutually exclusive competition between the γ and β genes for the LCR takes place to bring about multiple gene expression in single cells. Competition for the LCR also occurs in embryonic cells between the ε and γ genes (Wijgerde et al., 1995).

Of interest is the proposal that the LCR can activate the genes in the distal δβ when it is separated by over 30 kb of relatively closed chromatin. This indicates that the chromatin border which we have identified between the εγ and δβ domains is not an insulator which prevents enhancer-promoter interaction as has been shown for the chicken β globin HS:4 and the *Drosophila* scs/scs' and Su(Hw) elements. Instead the δβ chromatin border appears similar in function to the Fab-7 and Mcp elements of the *Drosophila* Bithorax complex (BX-C). These elements segregate the chromatin

of adjacent regulatory elements which control parasegment-specific expression of the *Abdominal-B* gene (*Abd-B*) (Mihaly et al., 1998) but do not block promoter-enhancer interactions.

Role of the LCR

Numerous transgenic experiments have shown that the LCR provides position independent, copy number dependent expression to linked genes. These data indicate that the LCR is sufficient to open every domain in which it integrates although it does not exclude the possibility that the LCR recruits or co-operates with endogenous elements to perform this function. Consistent activation of a LCR linked gene may reflect the fact that the gene now resides within the LCR domain. Clearly the human β gene is activatable in primitive cells when placed next to the LCR (Enver et al., 1990; Hanscombe et al., 1991) or in the position of the ε gene (Dillon et al., 1997) but not in its wild type location (Wijgerde et al., 1995) or when placed just 5' of the δ globin gene (Dillon et al., 1997).

Recent results have shown that the LCR is not required to maintain a nuclease sensitive chromatin structure in erythroid cells. This suggests that other elements in the locus are sufficient to perform this function. The fact that the globin locus is not a single large chromatin domain but actually consists of 3 domains raises the question, what activates and maintains the chromatin of these domains? The Δ2.5 transgenic results demonstrate that the region which we have shown to contain a chromatin border and developmentally specific initiator of domain transcription is required to properly activate the chromatin of the δβ domain in red cells. The transgenic data is supported by a naturally occurring 7.2 kb deletion (Corfu thalassemia) which encompasses the border and initiation region and part of the δ gene and results in inactivation of the intact β gene *in cis* (Galanello et al., 1990; Ribeiro et al., 1993). This shows that in the context of the full locus the LCR is not sufficient to activate the chromatin of the distal δβ domain. However the sequence elements within the 2.5 kb region alone are not enough to open this domain since the hispanic LCR deletion results in inactive chromatin throughout the entire locus (Forrester et al., 1990). Obviously the developmentally specific activation function provided by the 2.5 kb region must cooperate with the region deleted in the hispanic thalassemia to activate chromatin in the δβ domain. Once activation takes place local elements contained in the 2.5 kb region and within the domain may maintain active chromatin in the absence of the LCR.

Role of intergenic transcription

There are two possible explanations for the transcription of domains of DNase I sensitive chromatin in the human β globin locus *in vivo*. The first possibility is that it is simply the result of fortuitous transcription of active or accessible chromatin. The fact that the transcription is strand specific argues against this possibility. In addition we find that the δβ domain transcripts initiate from a single initiation site which correlates with the sharp transition in chromatin structure demarcating the 5' extent of the δβ domain (P. Fraser, unpublished). Though the results are preliminary it appears that the entire domain is transcribed from a precise initiation site in the border region. The fact that deletion of this initiation region in the Δ2.5 transgenics results in an absence of δβ domain transcripts and an activation failure of the genes in this domain suggests the second possibility, that intergenic transcription plays a role in the activation of the genes and/or chromatin.

It is becoming increasingly clear from the work done in other systems that transcription of non-coding RNA's is an essential component of gene regulation. In some cases such as the Xist RNA it is known that the non-coding RNA itself has a function. However the Xist locus also has another large non-coding transcript in the antisense direction which colocalizes on the active X chromosome. In this case and others, such as the immunoglobulin, IL4/IL13 and imprinted Igf2 and Igf2r genes it is not known whether the transcript or the transcription itself plays a role. It is possible that domain transcription in the human globin locus occurs as part of the mechanism of heritable marking of active genes or domains. It is known that a subset of RNA polymerase II complexes have chromatin remodelling activity and contain associated components such as SWI/SNF, histone acetyltransferases and cyclin dependent kinases (Cho et al., 1998) which may serve to modify chromatin in transcribed regions. Therefore we suggest that cell cycle dependent domain transcription may be a way of targeting entire domains for modification (Hebbes et al., 1994) rather than just gene local elements. This idea would fit in well with the observed transcription in early S phase around the time of replication of the globin locus, when newly replicated chromatin has to be assembled and modified. Alternatively, domain transcription may simply alter the topology of chromatin or displace inhibitory components such as linker histones resulting in a structure which is more accessible to trans factors. This type of activity would be beneficial in S phase and in early G1 phase as active chromosomal regions decondense after exiting mitosis.

Materials and methods

In situ hybridization analysis

Embryonic blood cells, fetal liver cells and anemic spleen cells were isolated form two different transgenic lines 15 and 72 (Strouboulis et al., 1992). Anemic mice were generated by 3 consecutive subcutaneous injections at 12 hour intervals of 0.4% phenylhydrazine hydrochloride dissolved in PBS (0.16ml/20g). Fetal livers and anemic spleens were disrupted into PBS by repeated pipetting as previously described (Wijgerde et al., 1995). BFU-E colonies were isolated after a 10 day culturing period of human cord blood progenitors in methyl cellulose.

Cells were fixed onto a poly-L-lysine coated slide (Sigma) 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. Slides were pre-treated for hybridization by a 0.01% pepsin digestion (4 min., 37°C) in 0.01M HCl, followed by a short wash in water and 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 moisturised chamber for 12hr.

For the detection of genic transcription the hybridization mixture contained 1ng/µl of four oligonucleotides (50nt) of β or γ globin intron sequences (Gribnau et al., 1998) containing a biotin side chain in the middle and on the 5' and the 3'end of the oligonucleotide (Eurogentec, Belgium). For the detection of intergenic transcription the hybridization mixture contained 1ng/µl of a digoxigenin-dUTP or a biotin-dUTP Nick translated probe. The following probe sequences according to the GenBank database were used for *in situ* hybridization analysis: a, ~-3439-18683; b, 3302-5208; c, 22663-25456; d, 30418-32403; e, 48910-50766; f, 52388-54776; g, 58041-61921.

Probes and oligonucleotides were dissolved in 50% formamide, 2× SSC, salmon sperm DNA (200ng/µl), 5x Denhardts, 1mM EDTA , 50mM sodium phosphate pH 7.0 and 50ng/µl mouse Cot DNA (Gibco BRL), denatured 5 minutes at 80°C, and renatured 15 minutes at 37°C. The coverslip was removed by dipping in 2x SSC and the cells were washed three times in 2× 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 amplification steps. PCNA was detected with an αPCNA antibody (DAKO) and a secondary antibody step with an antibody labeled with Texas-red. Slides were mounted with 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 250 cells were counted per data point.

BrdU labelling and detection

Fetal liver cells were isolated from 13.5 day transgenic embryos, containing a single integrated copy of the complete human β globin locus (Strouboulis et al., 1992). Fetal livers were disrupted as described and cells were incubated in 50µM BrdU for 15 minutes at 37°C under mild agitation. Cells were washed two times with PBS. Fixation, pre-treatment, hybridization and antibody detection were as described. After antibody detection of digoxigenin labeled probes cells were fixed for 1 hour in 3.7% paraformaldehyde/PBS. Slides were washed 10 minutes in PBS, dehydrated in 70%, 90% and 100% ethanol steps and air dried. DNA was denatured 2.5 minutes at 80°C in 50% formamide 2×SSC and washed 3 times 5 minutes in ice-cold 2×SSC. BrdU was detected with αBrdU antibodies (DAKO) and a secondary antibody layer with an antibody labeled with Texas-red.

Isolation of nuclei and general DNase I sensitivity studies

Nuclei were isolated from 10.5 day blood cells or adult anemic spleen cells as described by Forester *et al.* (1990). Cells were disrupted with 20 strokes of a type B pestle. Aliquots (50 µl) were digested with increasing amounts of DNase I for 3 minutes at 37°C. Reactions were treated with 250µg/ml proteinase K in 0.3 M NaCl, 0.5% SDS, 5mM EDTA and 10mM Tris pH 8 at 53°C overnight, phenol/chloroform extracted and ethanol precipitated. Pellets were dissolved in 50 µl H_2O. For Southern blot analysis DNA was digested with the appropriate restriction enzyme and 15-30µgr of DNA was loaded per lane. Embryonic series were digested with HindII+AlwNI and blots were hybridized with probes c, detecting a 2.2 kb AlwNI fragment; d, detecting a 4.9 kb HindII-AlwNI fragment; e, detecting a 1.5 kb HindII-AlwNI fragment; f, detecting a 2.6 kb HindII fragment; g, detecting a 1.8 kb HindII fragment and βIVSII (62663-63580) detecting a 2.2 kb HindII-AlwNI fragment. Adult series were digested with either HindII+AlwNI or EcoRI+SphI. Blots with EcoRI + SphI digests were hybridized with probes c, detecting a 4.2 kb EcoRI fragment; d, detecting a 7.0 kb EcoRI fragment; e, detecting a 5.1 kb SphI-EcoRI fragment; f, detecting a 1.4 kb SphI-EcoRI, a 1.7 kb EcoRI and a 2.3 kb EcoRI-SphI fragment; g, detecting a 3.5 kb EcoRI-SphI fragment and βIVSII detecting a 2.0 kb SphI-EcoRI fragment. For relative comparison of the series blots were hybridized with a Zfp probe (Mazarakis et al., 1996) detecting a 3.6 kb HindII-AlwNI fragment or a NF-M probe detecting a 3.6 kb HindII-AlwNI or a 1.9 kb EcoRI-SphI fragment. Probes were competed 15 min at 65°C with 10µgr/µl sonicated mouse genomic DNA in 5×SSC. The general sensitivity was quantitated via Phosphor imager analysis.

References

Ashe H.L., Monks J., Wijgerde M., Fraser P., Proudfoot N.J. 1997. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11**:2494-509.

Baptist M., Dumont J.E., Roger P.P. 1993. Demonstration of cell cycle kinetics in thyroid primary culture by immunostaining of proliferating cell nuclear antigen: differences in cyclic AMP-dependent and -independent mitogenic stimulations. *J Cell Sci* **105**:69-80.

Bernards R., Flavell R.A. 1980. Physical mapping of the globin gene deletion in hereditary persistence of foetal haemoglobin (HPFH). *Nucleic Acids Res* 8:1521-34.

Calzolari R., McMorrow T., Yannoutsos N., Langeveld A., Grosveld F. 1999. Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and deltabeta-thalassemia affects beta- but not gamma-globin gene expression. *EMBO J* **18**:949-58.

Cho H., Orphanides G., Sun X., Yang X.J., Ogryzko V., Lees E., Nakatani Y., Reinberg D. 1998. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol* **18**:5355-63.

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

Dillon N., Trimborn T., Strouboulis J., Fraser P., Grosveld F. 1997. The effect of distance on long-range chromatin interactions. *Mol Cell* 1:131-9.

Enver T., Raich N., Ebens A.J., Papayannopoulou T., Costantini F., Stamatoyannopoulos G. 1990. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* **344**:309-13.

Epner E., Reik A., Cimbora D., Telling A., Bender M.A., Fiering S., Enver T., Martin D.I., Kennedy M., Keller G., Groudine M. 1998. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell* 2:447-55.

Galanello., R., Melis, M.A., Podda., A., Monne, M., Perseu, L., Loudianos, G., Cao, A., Pirastu, M., and Piga, A. 1990. Deletion delta-thalassemia: the 7.2 kb deletion of Corfu delta beta-thalassemia in a non-beta-thalassemia chromosome. *Blood* **75**:1747-1749.

Gribnau J., de Boer E., Trimborn T., Wijgerde M., Milot E., Grosveld F., Fraser P. 1998. Chromatin interaction mechanism of transcriptional control in vivo. *EMBO J* 17:6020-7.

Grosveld F., van Assendelft G.B., Greaves D.R., Kollias G. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**:975-85.

Hanscombe O., Whyatt D., Fraser P., Yannoutsos N., Greaves D., Dillon N., Grosveld F. 1991. Importance of globin gene order for correct developmental expression. *Genes Dev* 5:1387-9.

Hebbes T.R., Clayton A.L., Thorne A.W., Crane-Robinson C. 1994. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J* **13**:1823-30.

Humbert C., Santisteban M.S., Usson Y., Robert-Nicoud M. 1992. Intranuclear co-location of newly replicated DNA and PCNA by simultaneous immunofluorescent labelling and confocal microscopy in MCF-7 cells. *J Cell Sci* **103**:97-103.

Li G., Lim K.C., Engel J.D., Bungert J. 1998. Individual LCR hypersensitive sites cooperate to generate an open chromatin domain spanning the human beta-globin locus. *Genes Cells* 3:415-29.

Long Q., Bengra C., Li C., Kutlar F., Tuan D. 1998. A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta-globin locus control region. *Genomics* 54:542-55.

Mazarakis N., Michalovich D., Karis A., Grosveld F., Galjart N. 1996. Zfp-37 is a member of the KRAB zinc finger gene family and is expressed in neurons of the developing and adult CNS. *Genomics* 33:247-57.

Mihaly J., Hogga I., Barges S., Galloni M., Mishra R.K., Hagstrom K., Muller M., Schedl P., Sipos L., Gausz J., Gyurkovics H., Karch F. 1998. Chromatin domain boundaries in the Bithorax complex. *Cell Mol Life Sci* 54:60-70.

Milot E., Strouboulis J., Trimborn T., Wijgerde M., de Boer E., Langeveld A., Tan-Un K., Vergeer W., Yannoutsos N., Grosveld F., Fraser P. 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* 87:105-14.

Nuez B., Michalovich D., Bygrave A., Ploemacher R., Grosveld F. 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* 375:316-8.

Perkins A.C., Sharpe A.H., Orkin S.H. 1995. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* 375:318-22.

Perkins A.C., Gaensler K.M., 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 U S A* 93:12267-71.

Raich N., Enver T., Nakamoto B., Josephson B., Papayannopoulou T., Stamatoyannopoulos G. 1990. Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science* 250:1147-9.

Reik A., Telling A., Zitnik G., Cimbora D., Epner E., Groudine M. 1998. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol* 18:5992-6000.

Reitman M., Lee E., Westphal H., Felsenfeld G. 1993. An enhancer/locus control region is not sufficient to open chromatin. *Mol Cell Biol* 13:3990-3998.

Ribeiro, M.L.S., Gu, L-H., Buchanan-Adair, I., and Huisman, T.H.J. 1993. Incorrect genetic counseling of a couple with beta-thalassemia, due to incomplete testing. *Am. J. Hum. Genet.* 52:842-843.

Southwood C.M., Downs K.M., Bieker J.J. 1996. Erythroid Kruppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev Dyn* 206:248-59.

Strouboulis J., Dillon N., Grosveld F. 1992. Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev* 6:1857-64.

Tewari R., Gillemans N., Wijgerde M., Nuez B., von Lindern M., Grosveld F., Philipsen S. 1998. Erythroid Kruppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region. *EMBO J* 17:2334-41.

Trimborn T., Gribnau J., Grosveld F., Fraser P. 1999. Mechanisms of developmental control of transcription in the murine alpha-and beta-globin loci. *Genes Dev* 13:112-24.

Wijgerde M., Grosveld F., Fraser P. 1995. Transcription complex stability and chromatin dynamics in vivo. *Nature* 377:209-13.

Wijgerde M., Gribnau J., Trimborn T., Nuez B., Philipsen S., Grosveld F., Fraser P. 1996. The role of EKLF in human beta-globin gene competition. *Genes Dev* 10:2894-902.

Wood W.I., Felsenfeld G. 1982. Chromatin structure of the chicken beta-globin gene region. Sensitivity to DNase I, micrococcal nuclease, and DNase II. *J Biol Chem* 257:7730-6.

Zafarana G., Raguz S., Pruzina S., Grosveld F. and Meyer D. 1996. The regulation of beta globin gene expression: the analysis of HS5 in the LCR. Haemoglobin switching. G. Stamatoyannopoulos (ed.) . Intercept Ltd. Hampshire, UK.

Chapter 7

General Discussion

Transcriptional activation of the β globin locus

For the last two decades the human β globin locus has provided a wealth of information concerning the regulation of multi-gene loci. The β globin locus consists of 5 functional genes, 5'ε-^{G}γ-^{A}γ-δ-β-3', that are developmentally expressed in the order of their organisation (Collins and Weissman 1984). Initial studies using transgenic mice to test the expression of small constructs comprising a single globin gene indicated that the essential information for high level expression was not present in these constructs (Magram 1985, Townes 1985, Kollias 1986, Chada 1986). Similar results were reported in a Dutch thalassemia patient, having intact β gene sequences without β gene transcription (van der Ploeg 1980). This information let to the discovery of a region, located upstream of the area containing the 5 functional genes, that is required for position independent and copy number dependent high level of globin gene transcription (Grosveld 1987). The region contains 5 erythroid specific developmentally stable hypersensitive sites and was designated LCR. Transgenic studies using full locus constructs with small deletions of individual hypersensitive sites of the LCR showed that proper activation by the LCR requires all hypersensitive sites *in cis* to overcome position effects (Milot 1996), suggesting that the different hypersensitive sites of the LCR act together as a holo-complex.

Three different models have been proposed explaining the mechanism of LCR activation and developmental regulation of the β globin genes; an accessibility, a tracking and a looping model. The accessibility model assumes that the LCR is only necessary for the initial activation of chromatin after which trans-acting factors are sufficient for proper developmental expression (Martin 1996). The tracking model explains LCR activation by considering the LCR as a nucleation site for the entry of trans-acting factors that track along the DNA in order to activate downstream genes (Herendeen 1992, Tuan 1992). In the looping model the LCR activates the locus and subsequently forms dynamic, direct chromatin interactions with gene-local regulatory elements (see Ptashne 1988 and Wijgerde 1995 and refs. therein).

In the absence of direct visualisation of chromatin dynamics *in vivo* it is only possible to use indirect measures in order to discriminate between the three models. An important parameter in the regulation of the β globin genes is that the different globin genes compete for the activation by the LCR. Competition for regulatory elements has been reported in many other multi-gene loci, like the Hox clusters (van der Hoeven 1996) and the Igf2-H19 genes (Webber 1998). Initial indications for competition in the β globin locus came from experiments using constructs with γ and β genes linked in different order to the LCR (Enver 1990, Hanscombe 1991). LCR-β constructs are expressed at all stages during development, however aberrant expression of the β gene in embryonic erythrocytes is ablated through insertion of a γ gene *in cis* between the β gene and the LCR.

More recently it was shown that addition of a marked β gene at different positions in the β globin locus results in a competitive advantage of the gene located closest to the LCR and that the relative distance between the genes and the LCR correlates directly with the expression levels (Dillon 1997). These results contradict a transcriptional interference mechanism that has been proposed to explain competition in an accessibility model. Transcriptional interference would predict that moving an additional β globin gene closer to the wildtype gene would result in a decrease in expression of the wildtype gene, in fact the opposite result was observed. The results reported by Dillon *et al.* (1997) are also inconsistent with a tracking model that only predicts changes in relative output levels when the linear order of genes is changed.

The data are consistent with a looping model with direct LCR-gene interactions in which the frequency and duration of interactions are the key parameters in determining the final output of the different genes. The frequency of interactions is dependent on the relative distance of the LCR to the gene whereas trans-acting factors are responsible for the duration of an interaction.

A pivotal difference between the looping model and both the accessibility and tracking models is that the looping model allows only one gene to be transcribed at a time whereas the other models propose co-initiation of multiple genes. Initial indications for the existence of co-activation came from an experiment showing that two γ promoters could be digested *in cis* using a restriction enzyme digestion assay in isolated nuclei (Bresnick and Felsenfeld 1994). The considerable amount of co-digested γ globin promoters were interpreted as being the consequence of simultaneous transcription initiation of both genes. However promoter hypersensitivity appears to reflect the potential to be activated, since HEL cells have hypersensitive δ and β promoters despite the absence of δ and β globin transcripts (Groudine 1983). Initial *in situ* hybridization studies also suggested the possibility of co-initiation of multiple genes *in cis*, since staining of fetal liver cells with antibodies against γ and β globin peptides resulted in many cells staining positive for both proteins (Fraser 1993). This was confirmed by single cell PCR studies (Furukawa 1995), and *in situ* hybridization experiments with intron specific oligo's also revealed loci with multiple gene signals *in cis* (Wijgerde 1995). The oligo's detect primary transcripts of actively transcribed genes and genes that were active within a period of 7 minutes as was determined by using the transcription inhibitor actinomycin-D. With these intron specific oligo sets γ and β globin primary transcripts could be detected *in cis* in a small subset of cells and could be an indication that both genes are co-initiated. Nevertheless the double signals could also be the result of a recent switch from γ to β globin expression or vice versa. In order to discriminate between these two possibilities we performed transcription inhibition and release studies using an elongation blocking agent DRB (chapter 4). We found a significant delay in the recovery of double signals *in cis* compared to double signals *in trans* after release of transcription, indicating that only a single β-like gene is activated at a time. Similar results were obtained with mini-γβ constructs (Ristaldi in prep.) and the murine β globin genes (chapter 5). Interestingly no delay in reappearance of double α and ζ globin signals was observed indicating that activation of the murine α genes is either more dynamic or that the different α genes are co-initiated *in cis*.

Chromatin domains in the globin locus

Indications for the presence of different chromosomal domains that could be involved in the regulation of the β globin locus were based on patient data (Collins and Weissman 1984). It was found that deletions encompassing the δ and β globin genes resulted in either a HPFH or a (δβ)° thalassemia phenotype and that the difference was delineated by a 700 bp region (Calzolari 1999 and refs. therein). A deletion of this region in transgenic mice results in an activation failure of the downstream β globin gene at fetal and adult stages (Calzolari 1999). Another deletion encompassing the ε silencer-enhancer showed that this region is important for ε and γ expression in the embryo (Liu 1997). Interestingly both regions seem to demarcate the borders of domains of active and inactive chromatin. We find that active globin genes reside in areas that are 2-3 fold more sensitive to DNase I

compared to regions containing inactive globin genes (chapter 6). DNase studies using blood cells of different developmental stages show that the human β globin locus is divided in at least three domains, a LCR, a εγ and a δβ domain. The LCR domain is in a constitutively open chromatin configuration in erythroid cells. The εγ domain is open in embryonic red cells and closes during fetal erythropoiesis, whereas the δβ domain is closed in the embryo, opens in the fetal liver and remains open in the adult. Similar results have been reported for the murine β globin locus using MEL cells. The active β^{maj} and β^{min} genes are located in a domain that is 3 fold more sensitive to DNase I than a domain that contains the inactive εy and βH1 genes (Smith 1984). The DNase I sensitive domain is depleted from histone H1 and differences in DNase I sensitivity are lost after unfolding higher order chromatin structures under low ionic strength conditions. Interestingly recondensation results in a restoration of differential sensitivities indicating that active domains are marked, which could be the result of general hyper-acetylation of histones in active domains, as has been found in the chicken locus (Hebbes 1994) or specific acetylation of certain lysine residues.

The Hispanic thalassemia deletion clearly shows that these 40 kilo bases are absolutely required for chromatin activation of the β globin locus (Forrester 1990). Numerous amounts of transgenic studies using LCR constructs comprising at least HS1-HS4 show that these hypersensitive sites are sufficient for chromatin activation of the β globin locus. Recently data has been published showing that the LCR is not required for the maintenance of DNase I sensitivity in the human and the murine β globin loci (Reik 1998, Epner 1998). The data indicate that information necessary for the maintenance of DNase I sensitivity is apparently retained outside the area defined as the LCR. Potential candidate elements involved in maintenance of an active chromatin configuration could be the elements like the ε silencer-enhancer, the 700 bp sequence 5' of the δ gene and several other elements including promoters and a LTR located upstream of HS5.

How do these local elements propagate DNase I sensitivity? Intriguing is the fact that active chromatin domains in the human β globin locus are marked by unidirectional domain transcripts (Ashe 1997, chapter 6). Similar results have been obtained in the heavy and light chain immunoglobulin loci were germ-line transcription precedes and is a prerequisite for VDJ recombination (for review see Willerford 1996). Germline transcription has been implicated to correspond with active chromatin, although it is not clear if active chromatin is required for proper recombination. Interestingly a deletion of a domain border in the β globin locus results in the absence of domain transcripts suggesting that these transcripts are initiated by elements that function as domain border. Similar start sites have been found in the LTR upstream of the LCR (Proudfoot pers. comm.), HS2 (Tuan 1992) and the ε silencer-enhancer (Allen 1983), and initiation of these transcripts could solely be fortuitous because of an active chromatin structure provided by local elements. Elements dispersed over the whole globin locus could propagate active and inactive chromatin domains by attracting PcG proteins, trxG proteins, chromatin remodelling complexes and/or histone acetyl transferases. Interestingly it has recently been shown that YY1 is a homologue of the *Drosophila* Polycomb group protein (PcG) pleiohomeotic. Several putative YY1 binding sites have been reported in the globin locus suggesting that PcG proteins and probably trxG proteins play a role in the maintenance of globin gene expression. Two other recent reports show that different chromatin remodelling complexes are involved in globin gene

expression (Armstrong 1998, O'Neill 1999). It remains however questionable how these elements can efficiently propagate active chromatin domains of more than 10 kb, since chromatin remodelling complexes in yeast seem to act only on a very localized area of a few nucleosomes (Kadosh 1998). Similarly cross-linking experiments and fluorescent staining with anti-bodies against PcG and trxG proteins suggest that these proteins do not spread but are associated with the immediate surroundings of PRE/TRE sequences (Pirrotta 1998, Orlando 1998). We found that domain transcripts are only present at certain stages of the cell cycle. If these transcripts would solely be the consequence of an active chromatin configuration then this implies that in certain stages in G1 and early S phase the chromatin structure is more accessible than in mid and late S phases and G2. These results and the uni-directionality of intergenic transcription rather suggest that domain transcripts or the polymerases that generate these transcripts play an active role in chromatin modification after mitosis and DNA replication. It has been shown that polymerases, that possibly produce domain transcripts, are associated with histone acetyl transferases or chromatin remodelling complexes. Transcription with these complexes will result in increased histone acetylation and the loss of H1, although a purely mechanic removal of H1 can not be excluded. It has indeed recently been shown that different mammalian RNA polymerase II complexes exist. One of these complexes contains the histone acetyl transferase protein pCAF associated with the elongating isoform of polymerase II (Cho 1998). This would suggest that the transcription process plays an active role in maintenance of an open chromatin structure. Another possibility is that the resulting RNA transcripts play a role in activation of the locus like has been postulated for the *Drosophila* male X chromosome. The male X chromosome is hyper-activated compared to the female X chromosomes. This hyper-activation is accompanied with the localization of roxA transcripts at the hyper-active X chromosome, and coating of the X chromosome has been implicated in the process of dosage compensation (Meller 1997).

EKLF and competition

Erythroid Kruppel-like factor plays an essential role in fetal and adult erythropoiesis as was shown by the generation of EKLF null mutants. Homozygous knock out mice suffer from severe anaemia and die around E12.5-E14.5 of gestation. Although EKLF$^{-/-}$ mice do not express the adult β globin gene it is not clear if the β globin genes are the only targets of EKLF. Recent studies using a μ-LCR-α-promoter-β^{maj} construct in order to rescue the fetal and adult β globin expression, indicate that more genes are affected by EKLF since no significant difference was observed between wildtype and transgenic mice in an EKLF$^{-/-}$ background (Wijgerde pers. comm.).

We studied the role of EKLF in competition of the β genes by generating compound transgenics with a single copy of the complete human β globin locus in an EKLF$^{+/-}$ and EKLF$^{-/-}$ background (chapter 3).We found that levels of EKLF are directly related to the ratio of β/γ globin expression, indicating that EKLF plays an important role in stabilising LCR-β gene interactions. A recent report shows that EKLF recruits a chromatin remodelling complex E-RC1 in order to form a hypersensitive site (Armstrong 1998). Interestingly a hypersensitive site could also be formed without the EKLF trans-activation domain, showing that chromatin remodelling and transcription initiation are independent processes that can clearly be separated.

Although these data suggest that EKLF is mainly required for the activation of the β globin promoter, DNase I sensitivity studies showed that HS3 of the LCR is also affected in EKLF$^{-/-}$ mice (chapter 3). EKLF is also a modulator of PEV, assayed in transgenic mice with a deletion of a single hypersensitive site of the LCR (Milot submitted). Increased amounts of EKLF reduce PEV of both γ and β globin expression whereas decreased amounts severely increase PEV, indicating that EKLF plays a much broader role in globin gene regulation than was previously thought. The results support a 'mass action' model for the LCR mediated activation of the globin locus, in which trans-acting factors have a certain probability of binding a target sequence (Grosveld 1999). Increased amounts of target sequences (like the several hypersensitive sites in the LCR) or increased levels of trans-acting factors result in a higher probability of activation of the locus.

Rather surprising was the finding that EKLF is already present and active in the embryonic red cells (Tewari 1998, chapter 5). Earlier reports show that a single β globin gene linked to an LCR is already active in the embryo (Enver 1990) and that expression is partially dependent on EKLF (Tewari 1998). Based on these results it was suggested that the β globin gene is silenced in the embryo as a consequence of competition from the ε and γ globin genes. However we do not detect a hypersensitive site at the β globin promoter in the embryo and find that the β gene is located in a domain of closed chromatin, despite the presence of active EKLF (chapter 6). Similarly, the relative frequency of LCR-β^M and LCR-β^{wt} interactions found by Dillon *et al.* (1997) in adult red cells, in a transgenic mouse line with the β^M gene replacing the ε gene, would predict activity of the wildtype β gene in the embryo. Nevertheless, no activity of the wildtype β gene could be measured at this stage, despite a 50% expression level of the β^M gene. This suggests that the β globin gene is inactive in the embryo in its natural environment and does not take part in the competition for the LCR. Interestingly the results indicate that EKLF can not bind the β globin promoter in inactive chromatin, and suggest that first an active chromatin configuration (probably 10 nm fibre) has to be established before EKLF can bind it's target sequence. Additional proteins that are not identified yet could be involved in the initial chromatin activation. However, one can not exclude the possibly that higher levels of active EKLF are sufficient to form a hypersensitive site. To date transgenic studies are in progress in which EKLF is linked to a mutated ligand binding domain of the estrogen receptor. Translocation of EKLF to the nucleus can be induced upon addition of a specific estrogen derivative. These mice could be used to test if increased levels of EKLF are sufficient to activate the β gene promoter in the embryo. Compound transgenics will be generated by crossing this line with a EKLF$^{-/-}$ line resulting in a EKLF 'knock on' mouse that can be very useful in monitoring the effects of induction of EKLF on chromatin structure and β gene activation.

References

Allan M., Lanyon W.G., Paul J. 1983. Multiple origins of transcription in the 4.5 Kb upstream of the epsilon-globin gene. *Cell* **35**:187-97.

Armstrong J.A., Bieker J.J., Emerson B.M. 1998. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* **95**:93-104.

Ashe H.L., Monks J., Wijgerde M., Fraser P., Proudfoot N.J. 1997. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11**:2494-509.

Bresnick E.H., Felsenfeld G. 1994. Dual promoter activation by the human beta-globin locus control region. *Proc Natl Acad Sci U S A* **91**:1314-7.

Calzolari R., McMorrow T., Yannoutsos N., Langeveld A., Grosveld F. 1999. Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and deltabeta-thalassemia affects beta- but not gamma-globin gene expression. *EMBO J* **18**:949-58.

Chada K., Magram J., Costantini F. 1986. An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature* **319**:685-9.

Cho H., Orphanides G., Sun X., Yang X.J., Ogryzko V., Lees E., Nakatani Y., Reinberg D. 1998. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol* **18**:5355-63.

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

Dillon N., Trimborn T., Strouboulis J., Fraser P., Grosveld F. 1997. The effect of distance on long-range chromatin interactions. *Mol Cell* **1**:131-9.

Enver T., Raich N., Ebens A.J., Papayannopoulou T., Costantini F., Stamatoyannopoulos G. 1990. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* **344**:309-13.

Epner E., Reik A., Cimbora D., Telling A., Bender M.A., Fiering S., Enver T., Martin D.I., Kennedy M., Keller G., Groudine M. 1998. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell* **2**:447-55.

Forrester W.C., Epner E., Driscoll M.C., Enver T., Brice M., Papayannopoulou T., 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**:1637-49.

Fraser P., Pruzina S., Antoniou M., Grosveld F. 1993. Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev* **7**:106-13.

Furukawa T., Navas P.A., Josephson B.M., Peterson K.R., Papayannopoulou T., Stamatoyannopoulos G. 1995. Coexpression of epsilon, G gamma and A gamma globin mRNA in embryonic red blood cells from a single copy beta-YAC transgenic mouse. *Blood Cells Mol Dis* **21**:168-78.

Grosveld F., van Assendelft G.B., Greaves D.R., Kollias G. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**:975-85.

Grosveld F. 1999. Activation by locus control regions? *Curr Opin Genet Dev* **9**:152-157.

Groudine M., Kohwi-Shigematsu T., Gelinas R., Stamatoyannopoulos G., Papayannopoulou T. 1983. Human fetal to adult hemoglobin switching: changes in chromatin structure of the beta-globin gene locus. *Proc Natl Acad Sci U S A* **80**:7551-5.

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

5:1387-94.
Hebbes T.R., Clayton A.L., Thorne A.W., Crane-Robinson C. 1994. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J* **13**:1823-30.
Herendeen D.R., Kassavetis G.A., Geiduschek E.P. 1992. A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science* **256**:1298-303.
Kadosh D., Struhl K. 1998. Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo. *Mol Cell Biol* **18**:5121-7.
Kollias G., Wrighton N., Hurst J., 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.
Liu Q., Bungert J., Engel J.D. 1997. Mutation of gene-proximal regulatory elements disrupts human epsilon-, gamma-, and beta-globin expression in yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A* **94**:169-74.
Magram J., Chada K., Costantini F. 1985. Developmental regulation of a cloned adult beta-globin gene in transgenic mice. *Nature* **315**:338-40.
Martin D.I., Fiering S., Groudine M. 1996. Regulation of beta-globin gene expression: straightening out the locus. *Curr Opin Genet Dev* **6**:488-95.
Meller V.H., Wu K.H., Roman G., Kuroda M.I., Davis R.L. 1997. RoX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* **88**:445-57
Milot E., Strouboulis J., Trimborn T., Wijgerde M., de Boer E., Langeveld A., Tan-Un K., Vergeer W., Yannoutsos N., Grosveld F., Fraser P. 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87**:105-14.
Milot submitted
O'Neill D., Yang J., Erdjument-Bromage H., Bornschlegel K., Tempst P., Bank A. 1999. Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. *Proc Natl Acad Sci U S A* **96**:349-54.
Nuez B., Michalovich D., Bygrave A., Ploemacher R., Grosveld F. 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* **375**:316-8.
Orlando V., Jane E.P., Chinwalla V., Harte P.J., Paro R. 1998. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J* **17**:5141-50.
Perkins A.C., Sharpe A.H., Orkin S.H. 1995. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* **375**:318-22.
Pirrotta V. 1998. Polycombing the genome: PcG, trxG, and chromatin silencing. *Cell* **93**:333-6.
Ptashne M. 1988. How eukaryotic transcriptional activators work. *Nature* **335**:683-9.
Reik A., Telling A., Zitnik G., Cimbora D., Epner E., Groudine M. 1998. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol* **18**:5992-6000.
Smith R.D., Yu J., Annunziato A., Seale R.L. 1984. beta-Globin gene family in murine erythroleukemia cells resides within two chromatin domains differing in higher order structure. *Biochemistry* **23**:2970-6.
Tewari R., Gillemans N., Wijgerde M., Nuez B., von Lindern M., Grosveld F.,

Philipsen S. 1998. Erythroid Kruppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region. *EMBO J* **17**:2334-41.
Townes T.M., Lingrel J.B., Chen H.Y., Brinster R.L., Palmiter R.D. 1985. Erythroid-specific expression of human beta-globin genes in transgenic mice. *EMBO J* **4**:1715-23.
Tuan D., Kong S., Hu K. 1992. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* **89**:11219-23.
Van der Hoeven F., Zakany J., Duboule D. 1996. Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls. *Cell* **85**:1025-35.
Van der Ploeg L.H., Konings A., Oort M., Roos D., Bernini L., Flavell R.A. 1980. Gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature* **283**:637-42.
Webber A.L., Ingram R.S., Levorse J.M., Tilghman S.M. 1998. Location of enhancers is essential for the imprinting of H19 and Igf2 genes. *Nature* **391**:711-5.
Wijgerde M., Grosveld F., Fraser P. 1995. Transcription complex stability and chromatin dynamics in vivo. *Nature* **377**:209-13.
Willerford D.M., Swat W., Alt F.W. 1996. Developmental regulation of V(D)J recombination and lymphocyte differentiation. *Curr Opin Genet Dev* **6**:603-9.

Summary

All our hereditary information is stored in DNA. Every human cell contains a copy of 3 billion base pairs of sequences organized in large units called chromosomes. The 23 human chromosome pairs contain about 100,000 genes. These genes code for different proteins which form the basis of all biological processes in our body. Different cell types are induced by activating different combinations of genes. One of these cell types is the red cell that originates from a stem cell via a differentiation process called erythropoiesis. During erythropoiesis stem cell specific genes are down regulated whereas red cell specific genes are activated. The majority of the cell mass of the resulting red cell consists of hemoglobin (Hb). Adult hemoglobin is a hetero tetrameric protein of two α and two β globin chains. Different α-like and β-like globin chains are expressed during development of the embryo, resulting in different hemoglobin tetramers. These different tetramers facilitate the oxygen uptake from the mother, since embryonic Hb has a higher affinity for oxygen than adult Hb.

The human β-like globin genes are organized in a locus that resides on chromosome 11. The locus contains 5 functional genes, 5'-ε-^{G}γ-^{A}γ-δ-β-3', which are organized according to their expression pattern during development. Two switches in the expression of the different β-like globin genes occur during development. The first switch from ε to γ globin expression takes place at the 6th week of gestation. Around birth γ globin expression is replaced by expression of adult δ and β globin genes. Five DNase I hypersensitive sites have been characterised in an area of 20 kilo bases just upstream of the ε globin gene. This locus control region (LCR) plays an important role in the regulation of the globin genes and is absolutely required for high level expression of the β-like globin genes. None of the hypersensitive sites appears to have developmentally specific gene activation capabilities. It has also been shown that the globin genes compete for the activation of the LCR, and competition is dependent on the relative distance of the gene from the LCR. These results suggest that the different hypersensitive sites act in concert, as a holocomplex, in order to activate downstream genes.

Three different models have been proposed to explain the mechanism of LCR mediated activation of the globin genes. The looping model predicts direct interactions between the LCR and the genes by looping out intervening DNA. In the accessibility model the LCR is only responsible for an open chromatin structure after which trans-acting factors activate the different globin genes. The tracking model explains activation of the globin genes by linear transport along the DNA of trans-acting factors initiated at the LCR. An important difference between the models is that the looping model allows only one gene to be activated at a time whereas both the accessibility and the tracking model allow multiple genes to be co-activated.

To address this question we applied an *in situ* hybridization technique which enables the detection of active genes and genes that have been active within a time span of 7 minutes before fixation. After *in situ* hybridisation these active genes light up as fluorescent signals within the cell. Co-detection of active γ and β globin genes in fetal liver cells of a transgenic mouse with the human β globin locus, results in a vast majority of single γ or β signals. Only a small number of loci have both γ and β gene signals. These double signals could be the consequence of a recent switch from γ to β gene activation or vice versa, but could also be the result of co-activation of multiple globin genes. In order to discriminate between these two possibilities we performed transcription inhibition and release experiments as has been described in

chapter 4 and 5. Polymerase II transcription elongation was reversibly blocked with DRB and the re-activation process was subsequently followed at different time points after the release of transcription elongation. The results show that single γ or β gene signals reappear twice as fast as double signals, whereas two signals on two different chromosomes reappear as fast as single signals. Similar results were obtained with the two human γ genes and the murine εy and βH1 genes and show that the human and murine LCR's can only activate one gene at a given moment. This is in agreement with the looping model and suggests that the LCR directly interacts with the globin gene promoters in order to activate these genes. The switching process is a highly dynamic process with the LCR forming semi stable interactions with a single gene promoter thereby switching back and forth between the active globin genes. The stability of LCR-promoter interaction and the frequency of these interactions are the key parameters that determine the final output of the different globin genes.

Erythroid krüppel like factor (EKLF) plays an important role in the activation of the adult β globin genes. Genetic ablation studies of EKLF have shown that this results in the absence of adult murine β globin genes. In chapter 3 we show that EKLF is also necessary for the expression of the adult human β globin gene and the formation of an erythroid specific hypersensitive site in the LCR. Reduction in the amount of EKLF results in a delayed switch from γ to β globin expression in mice transgenic for the full human β globin locus. These results suggest that EKLF plays an important role in the stabilisation of LCR-β globin promoter interaction.

Activation of the β globin locus not only accompanies the activation of the globin genes itself, but also results in transcription of non-coding areas located outside the globin gene sequences. In chapter 6 we studied this intergenic transcription and find that the globin locus can be divided in at least three different domains. LCR transcripts can be detected throughout development whereas transcription in the εγ domain is only present in embryonic blood and the early fetal liver. Transcription in the δβ domain is only detected in the fetal liver and in adult spleen cells. We find that areas with intergenic transcription have a higher sensitivity to DNase I indicating that these regions have an open/active chromatin structure. In adult blood cells the LCR is capable of activating the downstream δ and β genes without interference of an area of 30 kilobases of closed/inactive chromatin located in between. Since intergenic transcription is mainly unidirectional and is only present in G1 and early S phase of the cell cycle we suggest that intergenic transcription could play a direct role in the activation of chromatin.

Samenvatting

Al onze erfelijke informatie ligt opgeslagen in DNA. DNA is georganiseerd in lange strengen van in totaal 3 miljard opeenvolgende base paren. Elke lichaamscel bevat een kopie van dit erfelijk materiaal. Het DNA van een menselijke cel ligt verspreid over 23 chromosoomparen en bevat ongeveer 100.000 genen. De eiwitten waar deze genen voor coderen vormen de basis van alle biologische processen in ons lichaam. Door verschillende combinaties van genen te activeren kunnen verschillende celtypen geïnduceerd worden. Eén van die cel typen is de rode bloedcel, die onstaat door differentiatie van een stamcel. Dit differentiatie proces wordt erythropoiese genoemd en vindt plaats via verschillende tussenliggende 'precursor' cel stappen. Tijdens erythropoiese worden stamcel specifieke genen uitgeschakeld terwijl rode bloedcel specifieke genen, zoals de globine genen, worden geactiveerd. Rode bloedcellen bestaan voor het grootste deel uit hemoglobine (Hb), dat verantwoordelijk is voor het zuurstof transport in het lichaam. Hemoglobine is een hetero-tetrameer en bestaat uit twee α globine en twee β globine eiwitten. Tijdens de ontwikkeling van een embryo komen verschillende α en β globine eiwitten tot expressie, wat resulteert in verschillende Hb tetrameren. Deze verschillende tetrameren worden aangemaakt zodat het ontwikkelend embryo makkelijker zuurstof kan verkrijgen van de moeder. Embryonaal Hb heeft namelijk een hogere affiniteit voor zuurstof dan volwassen Hb.

De menselijke β globine genen liggen georganiseerd in een locus dat gesitueerd is op chromosoom 11. Het locus bestaat uit 5 functionele genen die georganiseerd zijn in de volgorde waarin ze tot expressie komen tijdens ontwikkeling (5'-ε-^{G}γ-^{A}γ-δ-β-3'). Tijdens ontwikkeling vinden twee omschakelingen in de expressie van de β globine genen plaats. De eerste omschakeling van ε expressie naar γ expressie vindt plaats rond de 6de week van de zwangerschap. Rond de geboorte vindt er een tweede omschakeling plaats van γ expressie naar expressie van het volwassen δ en β globine. In een gebied, ongeveer 20 kilobasen groot, dat gesitueerd is voor het ε globine gen zijn 5 DNase I hypergevoelige gebieden gekarakteriseerd. Elk DNase hypergevoelig gebied kan verschillende eiwitten binden en gebleken is dat deze gebieden een belangrijke rol spelen bij de activatie van de verschillende β globine genen. Verschillende experimenten hebben laten zien dat er geen DNase hypergevoelige gebieden van deze 'locus control region' (LCR) zijn die selectief één gen activeren of alleen maar actief zijn in een bepaalde periode tijdens de ontwikkeling van het embryo. Ook is aangetoond dat de verschillende actieve genen competeren voor activatie door de LCR en dat competitie afhankelijk is van de relatieve afstand van een gen tot de LCR. Dit suggereert dat de verschillende DNase hypergevoelige gebieden samenwerken, in de vorm van een holo-complex, om de verschillende globine genen te activeren.

Er zijn drie verschillende modellen gepostuleerd om de activatie van de globine genen door de LCR te verklaren. Het 'looping'-model voorspelt een directe interactie tussen de LCR en de verschillende globine genen, door tussenliggend DNA uit te 'loopen'. In het toegankelijkheids-model is de LCR alleen verantwoordelijk voor een open chromatine structuur waarna transcriptie-factoren de verschillende globine genen kunnen activeren. Het 'tracking'-model verklaart activatie van globine genen door lineair transport langs het DNA van, in de LCR geïnitieerde, transcriptie-factoren. Een belangrijk verschil tussen de drie modellen is dat het 'looping'-model slechts activatie van één gen toestaat op één bepaald moment, terwijl het toegankelijkheids- en het 'tracking'-model co-activatie voorspellen van actieve genen.

De door ons toegepaste *in situ* hybridisatie techniek maakt detectie mogelijk van actieve genen en genen die tot 7 minuten vóór fixatie actief zijn geweest. Na *in situ* hybridisatie lichten deze actieve genen op als kleine fluorescerende signalen in een cel. Co-detectie van actieve γ en β globine genen in foetale lever cellen van een transgene muis met het humane β globine locus laat voornamelijk loci zien met enkele γ of β globine signalen. Slechts een klein aantal loci laat zowel γ als β signalen zien. Deze dubbele signalen zouden een gevolg kunnen zijn van een recente switch in activatie van γ naar β globine of andersom, maar kan ook het gevolg zijn van co-activatie van beide genen in het zelfde locus. Om onderscheid te kunnen maken tussen deze twee mogelijkheden hebben we de in hoofdstuk 4 en 5 beschreven transcriptie inhibitie en re-activatie experimenten uitgevoerd. Hiertoe werd transcriptie elongatie van polymerase II reversibel geblokkeerd met DRB. Na uitwassen van DRB hebben we het transcriptie re-activatie proces gevolgd door middel van *in situ* hybridisatie met cellen geïsoleerd op verschillende tijdstippen na reactivatie. Deze experimenten laten zien dat enkele γ of β globine signalen twee keer sneller opkomen dan dubbele signalen, terwijl twee signalen op verschillende loci in dezelfde cel even snel opkomen als de enkele signalen. Hetzelfde resultaat werd gevonden wanneer de twee menselijke γ globine genen en de muis εy en βH1 genen werden gevolgd na reactivatie. Deze resultaten suggereren dat de LCR slechts één menselijk of muis β globine gen kan activeren op één bepaald moment. Dit is in overeenstemming met het 'looping'-model en suggereert dat voor de activatie van de globine genen de LCR een directe interactie aangaat met de promoters van deze genen. Tijdens het omschakelings-proces gaat de LCR dynamische directe interacties aan met de globine promoters en gaat daarbij heen en weer tussen de verschillende actieve genen. De stabiliteit en de frequentie van de LCR-promoter interacties zijn bepalend voor de mate van produktie van de verschillende globine genen.

Erythroid Kruppel like factor (EKLF) speelt een belangrijke rol in de activatie van de volwassen muizen β globine genen. Afwezigheid van EKLF resulteert in de afwezigheid van volwassen β globine. In hoofdstuk 3 laten we zien dat EKLF ook noodzakelijk is voor expressie van het humane volwassen β globine gen en de vorming van een DNase I hypergevoelig gebied van de LCR. Reductie van de hoeveelheid EKLF resulteert in een vertraagde switch van γ naar β expressie in transgene muizen met het humane β globine locus. Deze resultaten suggereren dat EKLF een belangrijke rol speelt bij de stabilisatie van LCR-β globine promoter interacties.

Activatie van het globine locus gaat niet alleen gepaard met activatie van de globine genen maar leidt ook tot transcriptie van niet coderende sequenties buiten de globine genen. In hoofdstuk 6 hebben we deze intergene transcriptie bestudeerd en vinden dat het menselijke β globine locus kan worden gesplitst in 3 verschillende domeinen. Transcriptie in de LCR wordt gedetecteerd op elk moment tijdens de ontwikkeling van het embryo, terwijl transcriptie in het εγ domein wordt gedetecteerd in het embryonaal bloed en in een vroeg stadium in de foetale lever. Transcriptie in het δβ domein daarentegen is alleen detecteerbaar in de foetale lever en in volwassen bloed cellen. Gebieden met intergene transcriptie hebben een hogere gevoeligheid voor DNase I wat betekent dat gebieden met intergene transcriptie een open/actieve chromatine structuur bezitten. In volwassen bloedcellen kan de LCR kennelijk de δ en β globine genen activeren zonder hinder te ondervinden van de aanwezigheid van ~30 kilobasen tussenliggend gesloten/inactief chromatine. Omdat intergene transcriptie unidirectioneel is en intergene transcriptie alleen wordt

waargenomen in G1 en vroege S fase van de cel cyclus zou intergene transcriptie een directe rol kunnen spelen bij de activatie van chromatine.

Curriculum Vitae

Joost Gribnau werd op 24 maart 1968 geboren te Hengelo (Ov.). In 1986 behaalde hij zijn VWO diploma aan het Twickel College te Hengelo. Aansluitend studeerde hij Biologie aan de Rijks Universiteit Leiden. Voor de afstudeerrichting Biochemie werd een doctoraal stage gelopen bij de vakgroep Moleculaire Plantkunde (Dr Mennes en Prof. dr Libbenga). Na het behalen van het doctoraal diploma Biologie, eind 1992 volgde een militaire dienstperiode als reserve officier bij de Dienst Militaire Gezondheidszorg. Tijdens deze dienstperiode voltooide hij het propedeuse Economie aan de Erasmus Universiteit Rotterdam. Na een korte stage bij Dagra Pharma trad hij eind 1994 in dienst als assistent in opleiding bij de afdeling Celbiologie en Genetica van de Erasmus Universiteit Rotterdam. Hier voerde hij onderzoek uit onder begeleiding van Prof. dr F.G. Grosveld en Dr P. Fraser resulterend in het hier gepresenteerde proefschrift.

Publications

A. Mennes, A. Quint, J. Gribnau, C. Boot, E. van der Zaal, A. Maan and K. Libbenga. 1992. Specific transcription and reinitiation of 2,4-D-induced genes in tobacco nuclei. *Plant Molecular Biology* 18: 109-117.

H. Bruins, J. Gribnau and D. Bwire. 1995. Investigation into typical and atypical tuberculin sensitivity in the Royal Netherlands Army, resulting in a more rational indication for isoniazid prophylaxis. *Tubercle Lung Disease* 76: 540-544.

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

F. Grosveld, E. de Boer, N. Dillon, J. Gribnau, T.McMorrow, E. Milot, T. Trimborn, M. Wijgerde and P. Fraser. 1998. The dynamics of globin gene expression and position effects. In: Epigenitics. Wiley, Chichester (Novartis Foundation Symp 214) pp 67-73.

F. Grosveld, E. de Boer, N. Dillon, J. Gribnau, E. Milot, T. Trimborn, M. Wijgerde and P. Fraser. 1998. The dynamics of globin gene expression and genetherapy vectors. *Seminars in Hematology* 35: 105-109.

P. Fraser, J. Gribnau and T. Trimborn. 1998. Mechanisms of developmental regulation in globin loci.*Current Opinion in Hematology* 5: 139-144.

J. Gribnau, E. de Boer, T. Trimborn, M. Wijgerde, E. Milot, F. Grosveld and P. Fraser. 1998. Chromatin interaction mechanisms of transcriptional control. *EMBO Journal* 20: 6020-6026.

T. Trimborn, J. Gribnau, F. Grosveld and P. Fraser. 1999. Mechanisms of developmental control of transcription in the murine α and β globin loci. *Genes & Development*, 13, 112-124.

Nawoord

Terugkijkend op de afgelopen vier jaar heb ik die tijd meer ervaren als hobby dan werk. Dit had veel te maken met de goede sfeer en collegialiteit op de 7e verdieping. Een speciaal woord van dank is hier zeker op zijn plaats. Frank, allereerst wil ik jou bedanken voor het prachtige wetenschappelijke klimaat dat jij wist te creëren en het gemak waarmee ik met vragen bij je terecht kon. Ik ben zelfs voetbal gaan waarderen in deze periode omdat ik moeilijk 'nee' kon antwoorden op je eerste vraag van mijn sollicitatie gesprek 'of ik graag voetbal?' Secondly I like to thank Peter for the time you spent on teaching me even the basic principles of molecular biology. I will definitely miss the many hours of discussions we had and the contagious determination with which you supervised the projects I was involved in. Good luck in Cambridge.

Natuurlijk wil ik alle mensen in lab 702 bedanken voor de gezellige chaotische sfeer. Tolleiv, een logeerbed stond altijd klaar als voetbal, kerstdiner of borrel weer eens had geleid tot bovenmatig drankgebruik. Zonder jou was dit verhaal er nooit geweest. Mark ik zal de congressen missen, die een combinatie waren van wetenschap, cultuur, goed eten en drinken en vele discussies. Ik hoop jullie veel te zien in the States. Eric we both couldn't resist the cantine food. Ernie wij hebben hoofdstuk 4 tot een mooi einde gebracht. Art, goed ingevallen en succes met schrijven. Karin we hoorden je graag komen en natuurlijk veel dank voor je hulp bij hoofdstuk 6. George (and Christina) you made us enjoy Greek food and drinks. Many thanks to Ali, Bas, Steffania, Mariken, Mariette, Gio, Rick, Dubi, John, Wouter, Edgar, Niels, Anna and the people that left the lab the recent years, Tara, Serafina, Roberta, Ton and Sara. An jij stond altijd klaar als er weer eens iets gelabeld moest worden, veel dank voor je in situ lessen. En niet te vergeten de rest van de globine groep op de 10de, met name mijn tweede begeleider Sjaak.

Melle en Rein bestellen was zo erg nog niet. Jan Jos en Ton Verkerk jullie gaven altijd thuis als er weer eens meer dan 2Mb aan elektronische post was binnen gekomen. Verder Marike en Rita, jullie waren altijd geïnteresseerd. En natuurlijk Mirko voor het snelle afhandelen van de vele spoed aanvragen. Tjeerd, Lien, Helen en Danielle veel dank ook namens mijn muizen voor de hulp en het geduld die wij vaak van jullie vroegen. Jopie, Joke, Wilma, Louise en Rob bedankt voor de vaat.

Ook veel vrienden buiten de universiteit hebben direct of indirect bijgedragen aan dit boekje, waaronder mijn paranimfen Henk Meijerink en Mike Klinkenberg. Henk met jou heb ik vele Leidse jaren met plezier de bench gedeeld. Vooral de tijd in de Hortus blijft ongeëvenaard. Mike de dienstplicht was zo erg nog niet en gaf je zelfs de kans eerder dan ik met hemoglobine in aanraking te komen.

Papa en Mama bedankt voor jullie onvoorwaardelijke steun en interesse de afgelopen jaren, mij ondertussen wijzend op de ethische aspecten van mijn werk. Oma 'Oldenzaal' bij jou is dit verhaal eigenlijk begonnen. De tochten door het lutterzand en andere mooie twentsche natuurgebieden op zoek naar goudvinken, ijsvogels en groene spechten hebben bij mij een blijvende interesse opgewekt voor de biologie. Als laatste en aller belangrijkste wil ik Susanne bedanken. Jij gaf me de steun en de rust die ik nodig had en begreep als beste waarom ik voor de wetenschap koos. Wij gaan er na de 30ste een mooie tijd van maken.