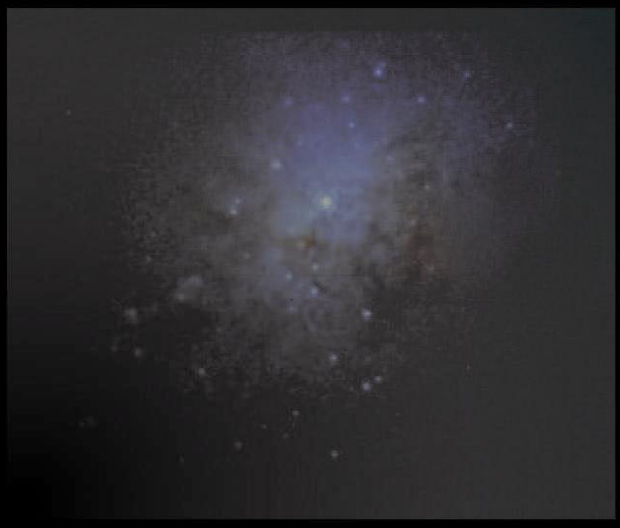
Globins in Space



Bas Tolhuis

September 2003

Globins in space

Globinen in de ruimte

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
Rector Magnificus

Prof.dr.ir. J.H. van Bemmel

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

woensdag 3 september 2003 om 9.45 uur

door

Bastiaan Johannes Tolhuis

geboren te Alkmaar

Promotiecommissie

Promotor:

Prof.dr. F.G. Grosveld

Overige leden:

Prof.dr. R. Kanaar
Dr. S.J.N. Philipsen
Dr.ir. D.N. Meijer

Co-promoter:

Dr. W.L. de Laat

Dit proefschrift is tot stand gekomen binnen de vakgroep Cel Biologie & Genetica binnen het Erasmus Medisch Centrum van de Erasmus Universiteit te Rotterdam. De vakgroep maakt deel uit van het Medisch Genetisch Centrum Zuid-West Nederland. Het onderzoek werd gesteund door NWO (ZonMw).

To bring the pieces back together, rediscover communication

M.J. Keenan

Voor Marjon, Joshua en Anna

Contents

Scope and outline of the thesis

The aim of this thesis was to gain a detailed understanding of transcriptional regulation in the context of the living cell. The β-globin locus was used as a model system to study the question: how distal *cis*-regulatory DNA elements communicate to activate tissue- and developmental stage-specific gene expression. This question was addressed through two different approaches. The first approach was based on the observation that globin gene transcription alternates between embryonic, fetal and adult genes in a flip-flop mechanism [111] and the assumption that direct contacts between the Locus Control Region (LCR) and gene promoters is required for transcription (looping model). To monitor these putative interactions in the living cell we distinctly tagged the human β-globin locus on both the LCR and the β-globin gene. This work is still in progress (chapter 2). In a second approach, we adopted Chromosome Conformation Capture (3C) technology [229] to measure the spatial conformation of the β-globin loci in man and mouse *in vivo* (see chapters 3 and 4). This demonstrated that spatial interactions between the LCR, actively transcribed genes and distal DNase I hypersensitive regions occur *in vivo*. Furthermore, the data support the existence of an erythroid cell-specific nuclear compartment dedicated to the transcription of the globin genes by RNA polymerase (RNAP) II, called the active chromatin hub (ACH). The ACH model provides a mechanistic framework that may improve knowledge of transcription in the 3-dimensional space of the nucleus and will be extensively discussed in the general discussion (chapter 5). The thesis starts with an introduction (chapter 1), discussing the key players involved in transcription, the nuclear architecture in relation to transcription, and β-globin gene transcription.

Chapter 1

Introduction

1.1 Summary

Many gene loci are regulated by an extensive organization of *cis*-regulatory DNA elements and an intricate system of cooperating *trans*-acting factors. The study of two of those loci, the β-globin loci of man and mouse is described in this thesis. Several *cis*-regulatory DNA elements that are distributed throughout the globin locus and a large number of *trans*-acting factors are essential for its proper transcriptional regulation. In the nucleus, chromatin is largely constrained in its motion, thereby limiting *cis*-regulatory DNA elements to a small volume, while *trans*-acting factors can diffuse "freely". Since the initial (bio)chemical reaction rates are proportional to the concentration of the molecules involved, the different dynamic behavior of *cis*-regulatory DNA elements and *trans*-acting factors raises the question of how they are able to reach a sufficiently high concentration to drive transcription. One possible strategy is to increase a local concentration of *trans*-acting factors in spatial proximity to the *cis*-regulatory DNA elements. Indeed, the nucleus has a number of separate compartments that all enclose transiently accumulated factors. The paradigm of such a nuclear compartment is the nucleolus, dedicated to the transcription of ribosomal RNA (rRNA) genes. It is a substructure that holds constrained chromatin as well as resident proteins that enter and exit this compartment at a high rate. This thesis describes data that provide insight into the regulation of the globin genes, which are transcribed by RNA polymerase (RNAP) II. Much like the nucleolus, the results described in this thesis suggest that there is a nuclear compartment consisting of clusters of *cis*-regulatory DNA elements, which in turn may mediate a transient accumulation of *trans*-acting factors. This nuclear compartment dedicated to RNAP II transcription of globin encoding genes has been named the "active chromatin hub" (ACH).

1.2 Transcription

Eukaryotes contain three different RNAP enzymes (RNAP I, II, and III) each dedicated to the transcription of a distinct set of genes. RNAP II transcribes mainly protein encoding genes and is the topic of this introduction. The transcription of a gene depends on several distinct features of RNAP II and the transcription machinery. First, it must be able to recognize gene-specific sequences in the bulk of DNA in order to initiate transcription properly. Second, the transcription of many genes takes place in a time and tissue-specific manner, while other genes are kept silent. Third, DNA in the nucleus is packaged into chromatin (figure 1.2.1). The ground state of native chromatin in eukaryotes is thought to be restrictive and not transcribed [1], while actively transcribed chromosomal domains have an altered chromatin structure (see box: DNase I) [2, 3]. Hence, there must be mechanisms to alleviate the restrictive forces. Two cooperating components are key in regulating RNAP II transcription. One component acts *in cis* and requires an element that is located on the same DNA molecule as the transcribed gene, the so-called *cis*-regulatory DNA elements. The other component acts *in trans*, i.e. it requires molecules separate from the gene containing DNA molecule, here referred to as *trans*-acting factors, which are a diverse group of proteins (and RNA) that act on the *cis*-regulatory elements and the transcribed gene.

Box: DNase I

Distortions of chromatin structure can often be detected by an increase in sensitivity to nucleases such as the endonuclease DNase I. This enzyme hydrolyzes double- or single-stranded DNA preferentially in a relative random manner at sites adjacent to pyrimidine nucleotides. Susceptibility to DNase I may therefore be used as an assay to measure the general accessibility of DNA regions in chromatin. Two types of DNase I sensitivity can be observed: general sensitivity and hypersensitivity. General sensitivity is often found in transcriptionally active chromosome domains. Hypersensitivity is restricted to small regions, often located at *cis*-regulatory DNA elements, within general sensitive domains. These so-called hypersensitive sites (HSs) are thought to reflect areas with a less dense nucleosomal packaging or nucleosome-free areas.

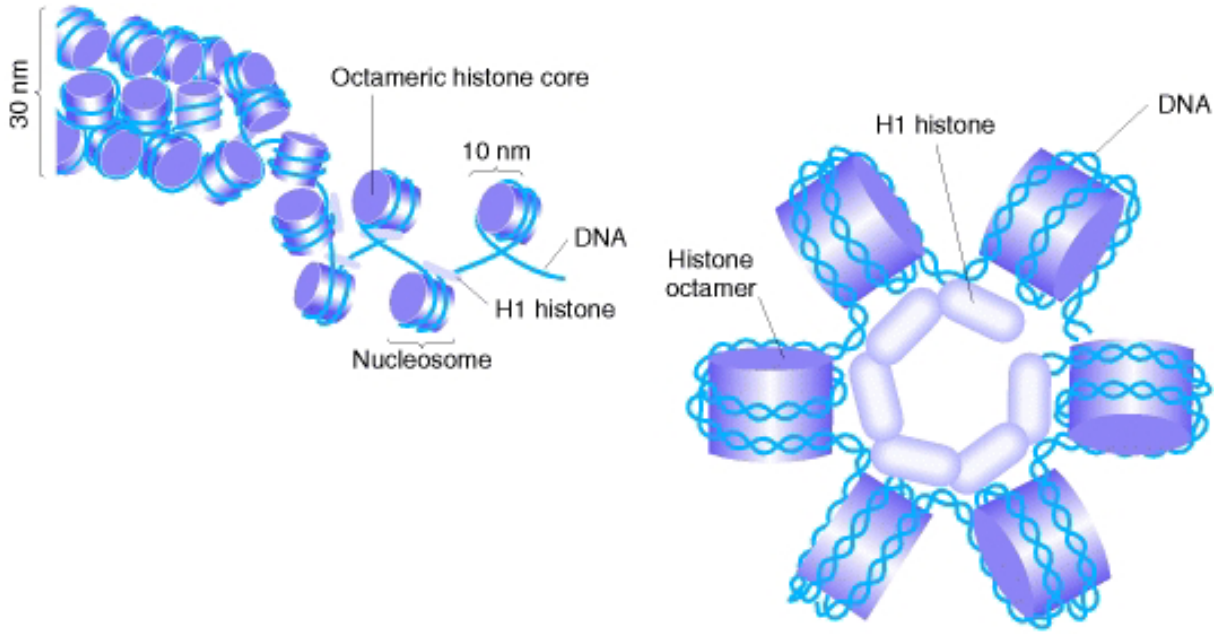


Figure 1.2.1. DNA packaged into chromatin by wrapping around nucleosomes.

The basic building block of chromatin is the nucleosome, which contains ~147 base pairs (bp) of DNA (blue fiber) wrapped 1.7-times around a core histone octamer (purple disk) in a 10 nm structure [288]. The core histone octamer consists of two copies of each histone H2A, H2B, H3, and H4. In the genome, thousands of nucleosomes are organized on a continuous DNA helix in strings separated by 10 to 60 bp of linker DNA. Each nucleosome can associate with histone H1 (grey/white rods) followed by coiling of the chromatin fiber into a solenoid structure with a diameter of 30 nm (on the left a side view and on the right top a view of the 30 nm chromatin fiber) (reviewed in [289]).

1.2.1 Cis-regulatory DNA elements

Promoters

The promoter specifies the site of transcriptional initiation *in vitro* and *in vivo.* The core RNAPII promoter is located approximately at -40 to +40 nucleotides relative to the RNA start site and contains at least one or a mixture of the following sequence elements, the TATA box, the TFIIB recognition element (BRE), the initiator (Inr), or the downstream promoter element (DPE) [4]. These sequences are sufficient to drive RNAP II transcription from a 'naked', i.e. not packaged into chromatin, DNA template *in vitro* [5]. From about -50 to -200 bp relative to the RNA start site there are typically multiple recognition sites for a subgroup of sequence-specific *trans*-acting proteins [4], here referred to as transcription factors. Together with the core promoter, these sequences can drive RNAP II in a reconstituted system [6] if the appropriate *trans*-acting factors are present. In mice, globin transgenes linked to their promoters are not transcribed to levels equal to that of endogenous genes [7]. *In vivo*, the promoters of active endogenous globin genes are DNase I HS [8].

Enhancers

Elements that increase transcription from a linked promoter are referred to as enhancers. Originally, they were defined in transient expression assays in cultured cells [9]. However, the definition does not specify the level of transcriptional enhancement or the assay used to measure it [10]. The enhancer effect is orientation-independent and shows some flexibility with respect to distance. In fact, distantly placed enhancers (> 1 kb) either upstream or downstream of a promoter are typical of higher eukaryotic genes [11]. Enhancers consist of DNA sequences that are recognized by transcription factors. The enhancers of the human β-globin gene can be detected as DNaseI HSs [8].

Enhancers, when linked to a promoter, seem to increase the probability of a gene establishing and maintaining an active transcriptional state in several experiments [11]. However, the presence of an enhancer is usually not sufficient to ensure high level of expression or expression in all cells that should express the gene when introduced into the mouse, or cultured cells, as a transgene. This is observed in various experiments. First, when a transgene is integrated stably into the genome of mammalian cell lines large differences in expression level are found, not related to presence or absence of an enhancer [12]. An increase in expression levels was found in mice carrying transgenes of β-globin constructs including its proximal enhancer elements when compared to β globin genes without the enhancers. However, expression levels did not correlate to the copy number of the integrated transgenes and were much lower than the expression levels of the endogenous genes [13, 14]. Similar results have been found for many genes and show that enhancers increase the proportion of the cells expressing a transgene, but cannot reproduce the normal expression pattern and/or level. In particular, they are often unable to overcome the restrictive effects of the chromatin structure at the site of integration of the transgene known as position effects. Interestingly, multiple copies inserts of a transgene carrying an enhancer can result in position-independent and copy-number dependent expression [15], suggesting that multimers of *cis*-regulatory DNA elements can cooperate to obtain proper levels of RNAP II transcription.

Locus Control Regions

Originally, Locus Control Regions (LCRs) were defined as dominant regulatory sequences that specify tissue specific, position-independent and copy-number dependent expression on a linked transgene in mice [16]. LCRs are capable of activating transcription of a single copy integrated transgene and mRNA levels per transgene copy are equal to mRNA levels of the endogenous gene [10]. Structurally, LCRs are composed of varying numbers of tissue-specific DNase I HSs and each site is composed of an array of sequences that can be bound by *trans*-acting factors [10, 17-19]. LCRs and the genes they regulate are often also separated by long stretches of intervening DNA. Furthermore, the location of individual HSs of a particular LCR can be spread out (figure 1.2.2) [10, 17, 20]. Many combinations of *cis*-regulatory DNA elements have been characterized in various vertebrate species that meet the functional definition of an LCR (reviewed [18]).

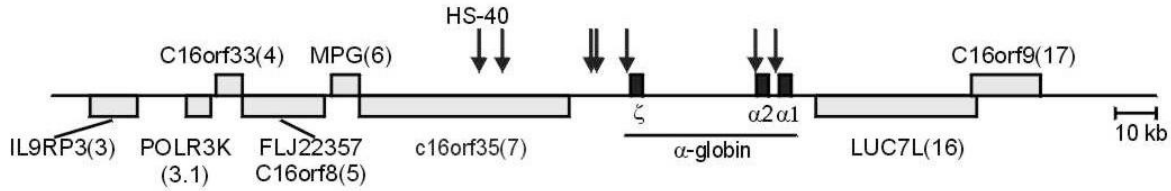
The most prominent property of LCRs is their strong, transcription-enhancing activity. When the LCR is absent, transcription of human β-globin transgenes is absent or very low [7, 21]. Deletion of the LCR from its endogenous position also eliminates or dramatically reduces transcription levels [22-26]. Another property of the LCR is that it can overcome a position effect induced mosaic expression pattern of a transgene, referred to as position effect variegation (PEV). Mice carrying a CD2 or β-globin transgene linked to their LCR show expression in all T cells or erythroid cells, respectively, even when it was integrated into the heterochromatic regions of the centromeres. In contrast, mice that lacked parts of the LCR show PEV [27, 28]. Importantly, the transgene constructs were DNase I insensitive in the nontranscribing portion of the cells [27, 28]. This indicates that the LCR operates by ensuring an open chromatin configuration that allows RNAP II transcription even in highly restrictive chromatin domains, such as (peri)centromeric heterochromatin (see below).

Insulators

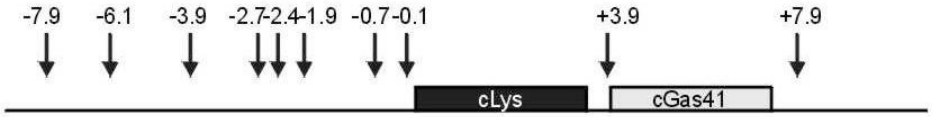
Insulators or boundary elements are DNA sequences defined operationally by two characteristics. They interfere with enhancer-promoter interactions when present between the enhancer and promoter. This can be experimentally defined in an enhancer-blocking assay.They are also thought to buffer transgenes from chromosomal position effects, which is experimentally measured by the ability to protect the gene in *cis* against PEV [29, 30]. The position of the insulator relative to the gene(s) and regulatory elements is essential to their function in that they must be placed between the anticipated source of silencing (e.g. chromatin) or activation (e.g. enhancer) and the affected gene locus (figure 1.2.3) [31]. Like other *cis*-regulatory DNA elements insulators are preferentially digested by nucleases [32, 33]. In addition, they contain consensus sequences for specific DNA binding proteins [34, 35].

A number of insulators have been reported in vertebrates (reviewed in [31, 36]) of which the chicken β-globin 5' HS4 (cHS4) is most relevant to our work [33, 37, 38]. The cHS4 was found to shield a reporter gene from the activating effects of a nearby mouse β-globin LCR in the human erythroleukemic cell line K562 [33]. Conversely it also protects against classical position effects in cell culture as HS4 protects the transcription of a stably transfected reporter gene in the absence of drug selection [38]. Most of the cHS4 insulating activity lies in a 250-bp CpG island (core element), which are often associated with promoters of housekeeping genes [37]. The human 5' HS5 was also found to have insulating properties in primitive cells, but not definitive cells [35]. In addition, a colony assay with 3' HS1 in beween the 5' HS2 enhancer and a reporter gene resulted in a 3-fold reduction of expression compared to a control construct without the 3' HS1. The authors interpreted this as an enhancer-blocking activity of 3' HS1 [39].

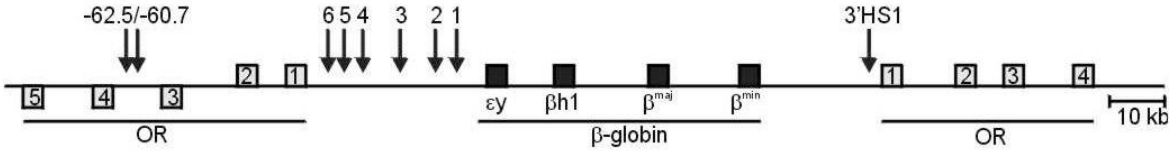
human α-globin cluster



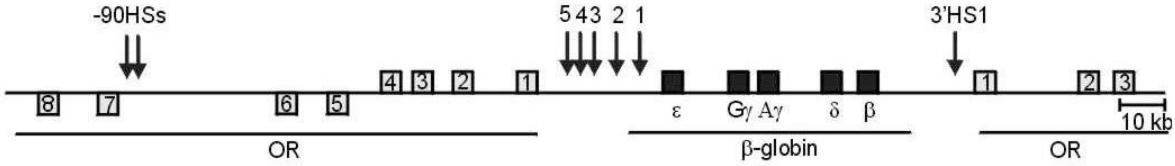
chicken lysozyme locus



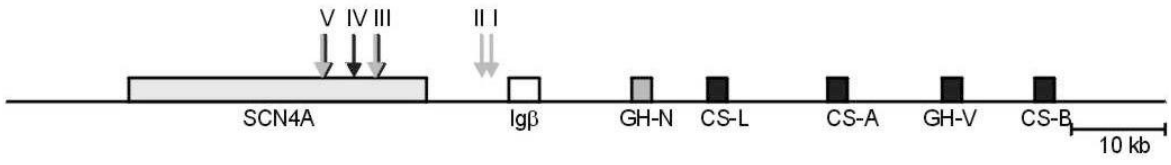
mouse β-globin cluster



human β-globin cluster



human Growth Hormone cluster



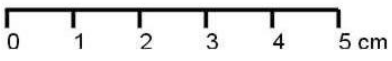


Figure 1.2.2. Genomic organization of several gene loci.

Boxes above the horizontal lines represent genes transcribing from left to right, boxes below represent genes transcribed in the opposite direction. Differentially expressed genes are indicated by different shades of grey and occur within one locus, but not separated by insulator or boundary elements. Arrows indicate known DNase I HSs (with grey/black colors corresponding to their target genes). The HSs are located throughout the depicted gene loci and often separated by a few kb to tens of kb. Note that the scale in kbs is indicated below each genomic fragment, except for the chicken lysozyme locus, where the numbering of HSs corresponds to the distance in kb from the transcriptional start site of the lysozyme gene. This figure was adapted from [20].

Synergy between elements

Despite their differences in functional definitions, there are two common features of *cis*-regulatory DNA elements: they can be detected as DNase I HSs *in vivo* and they are relatively small DNA fragments (200-300bp) that contain multiple binding sites for sequence-specific *trans*-acting factors. Those two properties suggest that *cis*-regulatory DNA elements represent regions that are bound by *trans*-acting factors. Clearly *cis*-regulatory systems somehow need to communicate with each other, in light of the fact that multiple *cis*-regulatory elements appear to be required to drive full transcription at endogenous loci and that combinations can overcome position effects in transgenic experiments.

1.2.2 *Trans*-acting factors

The transcription process starts with the assembly of a multiprotein complex at the promoter, the so-called preinitiation complex (PIC). Approximately 70 *trans*-acting factors are part of this preinitiation complex and required for its assembly [5]. Each of those factors contributes to the complexity of the initiation of RNAP II transcription. Transcriptional initiation culminates in the assembly of the general transcription machinery at the promoter. The key components are TATA-binding protein (TBP) and RNAP II.

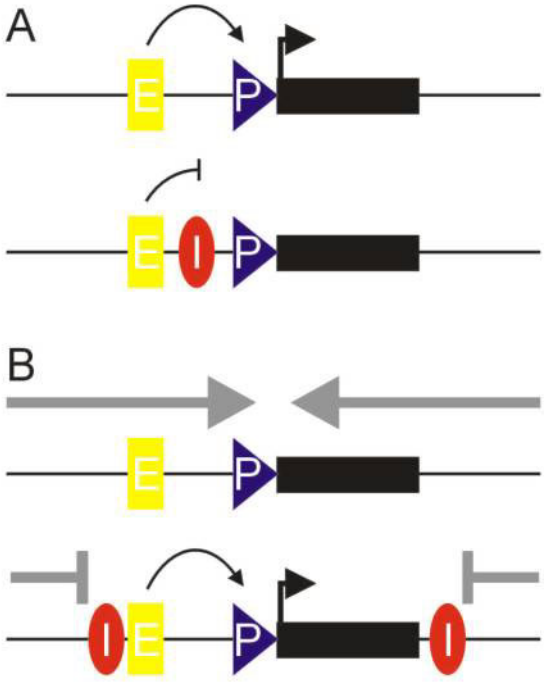


Figure 1.2.3. The position of an insulator relative to a gene is essential for proper functioning.

An enhancer (E, yellow box) can activate transcription (arrow) from a promoter (P, blue triangle) of a gene linked *in cis* (black rectangle) (A, top). In contrast, the enhancer activities are blocked when an insulator element (I, red oval) is located between the enhancer and promoter (A, bottom). Heterochromatin (grey arrows) spreads into the locus inhibiting transcription (B, top), while two flanking insulator elements act as a boundary and stop the heterochromatin spreading, consequently transcription takes place (B, bottom).

RNA polymerase II

RNAP II catalyzes DNA-dependent synthesis of RNA. It is a ~550 kDa protein complex composed of 12 subunits whose sequences are conserved among many diverse species. The largest subunit of RNAP II has a carboxy terminal domain (CTD), which is specific to RNAP II and plays key roles in the regulation of transcription initiation and coordination of co-transcriptional RNA processing events [40]. The CTD consists of a heptapeptide repeat with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The CTD may be long, e.g. it consists of 26 repeats in yeast and 52 in humans, and therefore extends far from the main body of RNAP II. During PIC assembly the CTD is hypophosphorylated, however, after transcription initiation the CTD is rapidly phosphorylated [41]. Although RNAP II catalyzes transcription of protein encoding genes, it is unable to initiate promoter-dependent transcription in the absence of other factors [5].

General transcription factors

Direct basal level of *in vitro* RNAP II transcription from a 'naked' DNA template, containing a promoter sequence, is facilitated by a group of proteins called the general transcription factors (GTFs). *In vitro* studies have revealed a stepwise recruitment of GTFs and RNAP II at the promoter to complete PIC assembly. This process commences with a single polypeptide that binds the TATA box of the promoter; this is the TATA-binding protein (TBP). Subsequent to formation of the TBP-DNA complex transcription factor II B (TFIIB) enters the PIC and its C-terminal domain binds the promoter through the BRE downstream of the TATA box. The binding by TFIIB is likely to account for the direction of transcription. Next, RNAP II associated to TFIIF enters the DNA-TBP-TFIIB ternary complex, which has a stabilizing effect. Binding of the RNAP II-TFIIF complex at the promoter allows entry of the next two GTF complexes, TFIIE and TFIIH. First, TFIIE is recruited, which affects late events in PIC assembly, including TFIIH recruitment and regulation. TFIIH is a large and complex GTF, consisting of 9 subunits and it is the only GTF with defined enzymatic activities. Two of its subunits are ATP-dependent DNA helicases of opposite polarity (XPB and XPD), which also have a role in nucleotide excision repair (NER) [290-292]. Another subunit is a cyclin-dependent protein kinase (cdk7-cyclin H), which is essential for transcription intitiation while it is dispensable for NER. Both TFIIE and TFIIH are required for ATP-dependent formation of the open promoter complex prior to initiation of transcription (reviewed in [5, 40]). However, this stepwise assembly only holds true for assembly of the transcription machinery *in vitro*. *In vivo* several reports show that the order of events may differ per gene locus rather than a 'general' order of recruitment [42-48]. Furthermore, in reconstituted systems and *in vivo* chromatin, additional factors are required to initiate transcription, these include several types of co-factors and sequence-specific transcription factors.

Co-factors

Two groups of co-factors have been described (although some can be placed in both categories): protein components of, or interacting with the transcription machinery, and modifiers of chromatin structure. First, TBP-associated factors (TAFs) form a complex with TBP, called TFIID, which was originally defined as a general transcription factor. The hallmark of TFIID is its ability to support activated transcription in reconstituted systems. TAFs can serve as core promoter recognition factors by binding to the Inr and DPE sequences, which are located downstream of the TATA box. This stabilizes the TFIID on the core promoter, enhancing or replacing TBP-TATA box interactions. Furthermore, certain TAFs are also co-factors that are capable of binding to activation domains of sequence-specific transcription factors and exhibit enzymatic activities that are capable of modifying

proteins and potentially modulating their activities (reviewed in [5]). Another co-factor complex that has been described is Mediator, a modular complex that 'mediates' interactions between the activation and repression domains of TFs and the CTD of RNAP II. Thus, it transduces both positive and negative regulatory information from TFs to the basal transcription machinery. Mediator is composed of approximately 25 proteins and is unable to bind specific DNA sequences. To date, the mechanism of Mediator action is unknown (reviewed in [6, 40]).

The chromatin modifying co-factors can be divided into two distinct sets based on their enzymatic activities. First, there are co-factors that covalently modify proteins by either adding chemical compounds to or removing them from other proteins. Second, there are co-factors that use the energy of ATP hydrolysis in order to perturb or reorganize chromatin structure. These are collectively referred to as chromatin remodeling enzymes.

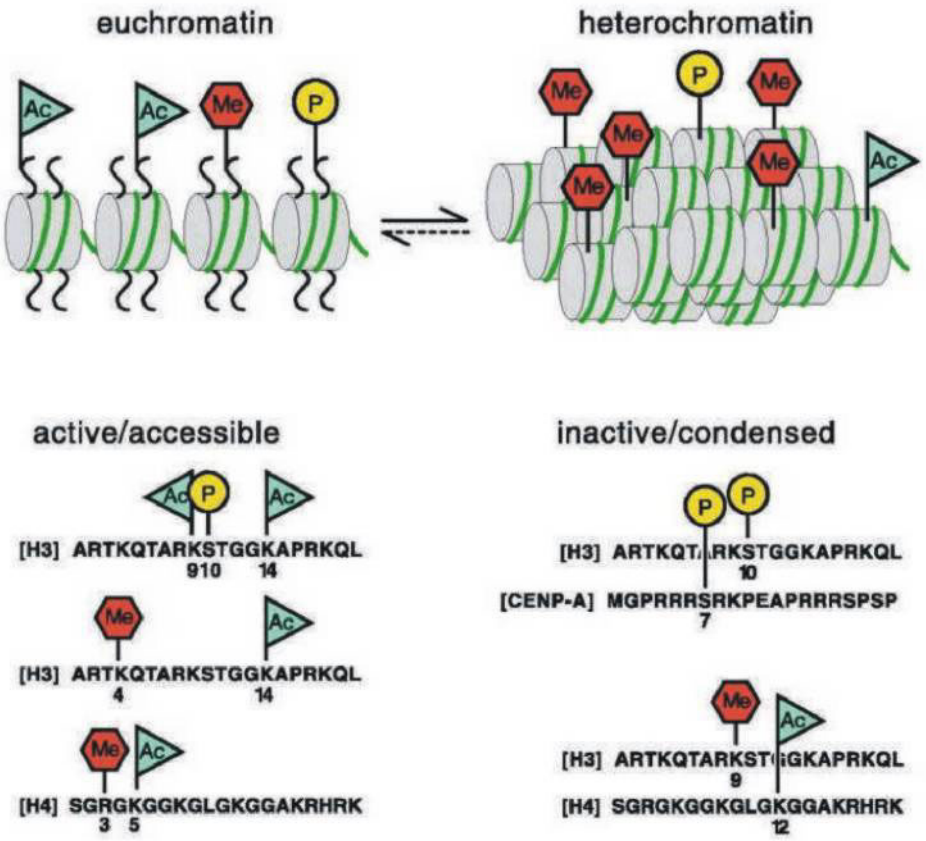


Figure 1.2.4. The 'histone code': models for euchromatic or heterochromatic histone tail modifications.

Top panel shows a schematic representation of euchromatin and heterochromatin as accessible or condensed nucleosome fibers containing acetylated (Ac, green triangles), phosphorylated (P, yellow circles), and methylated (Me, red hexagonals) histone NH_2-termini. Bottom panel illustrates examples of combinatorial modifications in histone NH_2-termini that are likely to represent "imprints" for active or inactive chromatin. Single-letter abbreviations for amino acid residues: A, Ala; E, Glu; G, Gly; H, His; K, Lys; L, Leu; M, Met; P, Pro; Q, Gln; R, Arg; S, Ser; and T, Thr. This figure was adapted from [53].

Co-factors that covalently modify proteins can alter the histones by a collection of post-translational modifications [49]. These modifications take place on the 'tail' domains of histones, which extend out of their globular domain, and some are known to correlate with nuclear processes (figure 1.2.4). For instance, histone acetyl transferases (HATs) function enzymetically by transferring an acetyl group from acetyl-coenzyme A to the ε-amino group of certain lysine residues within the histone tail [50]. This can result in histone hyperacetylation, which correlates with increased transcription [51]. The counterparts of HATs are histone deacetylases (HDACs) that remove the acetyl groups that leads to hypoacetylation, which correlates with transcriptional repression [51]. Histone methyl transferases (HMTases) target either lysine or arginine residues with different outcomes on transcriptional state. Methylation of lysine 9 and lysine 27 of histone H3 (H3-meK9 and H3-meK27, respectively) will lead to silencing and subsequent heterochromatinization. In contrast, H3-meK4 and arginine methylation at several residues of H3 and H4 correlates with an active state of transcription [52]. In the past, two hypothetical mechanisms have been proposed that explain these correlations between modification and nuclear processes. First, histone tails bind the DNA through charge interactions and modifications may alter chromatin structure by influencing histone-DNA and histone-histone contacts. This may lead to positive and negative changes in chromatin accessibility for other *trans*-acting factors. Second, a histone 'language' may be encoded on the tail domains. This 'language' is named the 'histone code' and may be read by other factors that are the actual agents of activity [49]. In fact, there are two protein domains found that associate with specific histone modifications. The bromodomain can selectively interact with acetylated lysines and chromodomains appear to be targeting modules for methylation marks [53]. Obviously, these two mechanisms may not be mutually exclusive but could very well co-exist *in vivo*.

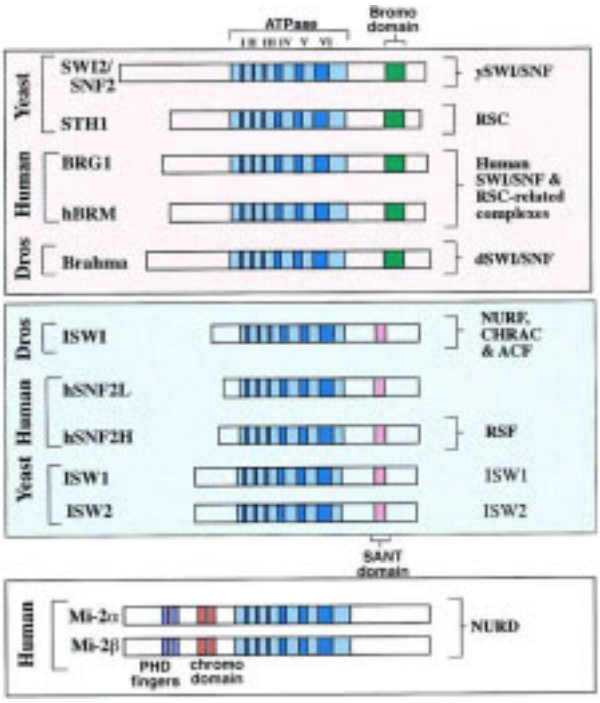


Figure 1.2.5 . ATPase subunits of chromatin remodeling complexes.

Conserved domains are labeled. Protein names are shown at left; names of complexes at right. SWI/SNF family complexes are depicted in the pink box (top), ISWI-family complexes in the blue box (middle) and Mi-2 complexes in the white box (bottom). This figure was adapted from [61].

Some HATs have been shown to acetylate nonhistone substrates as well as histones In fact, several erythroid-specific transcription factors are substrates for their HAT activities. GATA-1 can be acetylated *in vitro* and *in vivo* by p300/CBP and increases GATA-1-dependent transcription [54]. *In vivo* an acetylated form of erythroid kruppel like factor (EKLF) exists and interacts with PCAF and p300/CBP. However, *in vitro* it is only an acetylation substrate for p300/CBP [55]. The p300/CBP interacts with NF-E2, resulting in increased p300/CBP nucleosomal HAT activity and acetylation of NF-E2 [56]. These interactions suggest that posttranslational modifications can influence transcription factor activity and recruitment of HATs to specific *cis*-regulatory DNA elements.

Chromatin remodeling co-factors can perturb or reorganize chromatin structure in an ATP-dependent manner and this activity might relieve the restrictive and transcriptionally silent ground state. 'Chromatin remodeling' refers to numerous *in vitro* ATP-dependent changes in a chromatin substrate [57]. The *in vivo* activities of chromatin remodeling complexes are poorly understood. Only a few examples of *in vivo* chromatin remodeling activities have been reported [46, 58, 59]. Remodeling complexes are classified into three groups (figure 1.2.5) based on their biochemical properties and the overall sequence similarity of their ATPase units. First, the SWI/SNF-family of complexes mostly play a role in the activation of transcription. Second, ISWI members appear dedicated to transcriptional repression pathways and nucleosome assembly. Third, Mi-2/CHD complexes are also involved in transcriptional repression and they usually form complexes with HDACs [60, 61].

Some sequence-specific transcription factors interact with chromatin remodeling factors [62, 63]. For instance, two subunits of a SWI/SNF complex are necessary and sufficient for targeted chromatin remodeling and transcriptional activation by EKLF *in vitro*. They interact directly to generate a DNase I HS within the chromatin-assembled β-globin promoter. In contrast, these two subunits do not interact or function on the same promoter with two unrelated transcription factors [64]. In addition, SWI/SNF complexes contain a bromodomain, which has been shown to selectively interact with acetylated lysines [53, 61], suggesting an alternative targeting pathway mediated indirectly by HATs. Furthermore, human SWI/SNF complexes are phosphorylated during mitosis, which correlates with removal of the complexes from condensing chromosomes. Recent studies also suggest the existence of 'shielding' factors, which block ATP-dependent remodeling of chromatin fibers [65, 66]. This suggests that chromatin remodeling factors might be tightly regulated and may interact with other *trans*-acting factors, which would result in a proper targeting and timing of their activation or repressive activities.

Sequence-specific transcription factors

Transcription factors can be defined as proteins that recognize and bind to specific sequences that are usually located in *cis*-regulatory DNA elements and regulate transcription of specific genes through interactions with other factors. They contain a DNA-binding domain and one or more regulatory domains [50]. Transcription factors comprise a huge number of highly diverse proteins that can be expressed ubiquitously, tissue-specifically, or in a cell cycle-dependent manner. Describing all the members of this group is beyond the scope of this introduction. Here some general properties are recapitulated. First, several types of post-translational modifications mediated by other proteins can modulate the regulatory properties of certain transcription factors. For instance, the erythroid cell specific factors NF-E2, GATA-1, and EKLF can be phosphorylated [67-69], and acetylated [54-56]. Second, transcription factors may be involved in many protein-protein interactions, including those with the basal transcription machinery, other transcription factors and chromatin modifying co-factors. In this context, NF-E2 may modulate transcription through direct interaction with the basal transcription apparatus component $TAF_{II}130$ [70] and EKLF needs a specific chromatin remodeling complex, E-RC1, to regulate transcription [63, 64]. Recently, GATA-1 was found to interact with several proteins *in*

vivo using a single step purification method ([71] and Patrick Rodriguez, personal comm.), some of these are thought to be activating and others repressing transcription. Third, some factors can bind chromatin independently, while others require separate co-factors to facilitate binding in reconstituted systems. NF-E2 binding results in an ATP-dependent disruption of nucleosome structure, suggesting that chromatin remodeling factors present in the *Drosophila* embryonic assembly extract may be involved [72, 73]. EKLF requires the E-RC1 complex as separate co-factor to bind chromatin [63]. Recently, it has been shown that transcription factors HNF3 and GATA-4 can bind their specific binding-sequence without the activities of chromatin remodeling factors. [74]. The authors speculate that the ability of GATA-4 to bind and open chromatin may relate to the ability of the erythroid cell specific factor GATA-1 to disrupt a subset of histone-DNA contacts on a mononucleosome [54]. The highly conserved zinc finger motif of this family of factors may generally possess a chromatin disrupting function. The sequence-specific chromatin binding activity and the fact that transcription factors are involved in many protein-protein interactions provides a mechanistic framework for recruitment of other *trans*-acting factors at any given *cis*-regulatory DNA element.

How do *cis*-regulatory elements and *trans*-acting factors collaborate efficiently on the restrictive chromatin template to assemble the preinitiation complex at the promoter, which is followed by RNAP II transcription? Each *cis*-regalutory DNA element by itself appears to be a target of certain sequence-specific transcription factors and their interacting partners. However, many gene loci have widely separated regulatory elements that all appear to contribute to proper transcriptional regulation. One important aspect of (bio)chemistry is that the initial rate of chemical reactions is proportional to the concentration of molecules involved. An economical way to bring the different components together might be through compartmentalization [75].

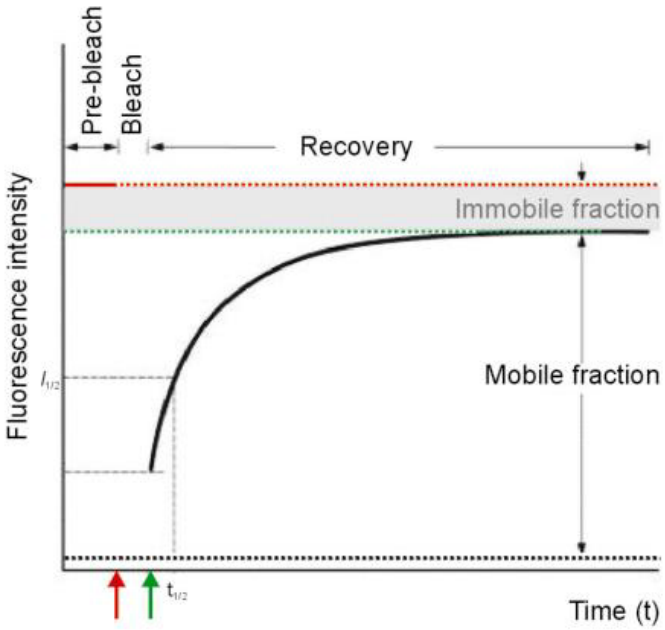


Figure 1.3.1. Quantitative fluorescence recovery after photobleaching (FRAP).

An idealized plot of fluorescence intensity as a function of time shows the parameters of a quantitative FRAP experiment. The bleach region is monitored during a pre-bleach period to determine the initial intensity (red solid/dotted line). This region is bleached using high-intensity illumination (red arrow), and recovery is monitored starting at t_0 (green arrow) until the fluorescence intensity reaches a final value, when no further increase can be detected (green dotted line). Some methods calculate the effective diffusion coefficient, D_{eff}, directly from the time ($t_{1/2}$) to reach half final intensity ($I_{1/2}$). To calculate D_{eff} as accurately as possible, the fluorescence intensity must be corrected for the background intensity (black dotted line) and the amount of total fluorescence removed by the bleach. The mobile fraction is the proportion of fluorescence that is regained, indicated by the difference between the initial and final fluorescence (grey area). This figure was adapted from [76].

1.3 The nucleus

Eukaryotic cells contain a specialized cell organelle that holds the genetic material (DNA), referred to as the nucleus. Based on microscopic studies, it has been suggested that the nucleus is organized into many distinct compartments. These so-called nuclear compartments were originally defined by the following parameters: they are morphologically identifiable structures using either light or electron microscopy, and contain a particular subset of resident proteins. In addition, some compartments can be biochemically isolated in an enriched form, although they lack any delineating membranes. In addition, kinetic studies (see box: FRAP and FLIP) demonstrate that the nucleus can also be subdivided into highly dynamic components and sections with constrained mobility.

Box: FRAP and FLIP

Quantitative fluorescence recovery after photobleaching (FRAP) generates information about the relative mobility of the fluorescently (e.g. GFP) tagged protein: the effective diffusion coefficient (D_{eff}). In FRAP a region is bleached irreversibly and then recovery of fluorescence in the bleached zone is monitored. To derive D_{eff}, the recovery is plotted as a function of time (figure 1.3.1). By comparing pre-bleach fluorescence intensity and the recovered intensity the protein can be separated in mobile and immobile fractions. Proteins are considered to have a high mobility, if they have both a high D_{eff} (i.e. fast diffusion rate) and a high mobile fraction. Studies based on fluorescence loss in photobleaching (FLIP) show continuity or transport between different populations of proteins, e.g. when fluorescently tagged proteins occupy distinct nuclear compartments. In FLIP, a region is repeatedly bleached, and the loss of fluorescence from outside the bleached zone is monitored. The rate of loss of fluorescence signal is an indicator of the mobility or exchange rate of the protein. The fluorescence signal from freely mobile molecules decreases rapidly, the fluorescence signal from slowly mobile molecules decreases slowly. Note that in this method the movement of unbleached molecules is monitored excluding the possibility that damage by the bleach pulse affects the mobility of the monitored proteins.
(For details see [76, 77])

1.3.1 The nuclear envelope

The nucleoplasm (i.e. the interior of the nucleus) is physically separated from the cytoplasm by a set of membranes and an associated protein network, collectively labelled the nuclear envelope (NE). Structurally the membrane components are an outer nuclear membrane (ONM) that faces the cytoplasm, and an inner nuclear membrane (INM) that faces the nucleoplasm. The INM has associated to it a unique set of integral membrane proteins, e.g. Lamin B receptor (LBR), lamina associated polypeptides (LAPs), and emerin. Most of those proteins interact with a network of polymers formed by lamins (i.e. nuclear intermediate filament proteins) called the nuclear lamina that is closely associated with the INM. Due to its polymeric nature, the nuclear lamina provides a scaffolding structure of the NE (reviewed in [78]).

Enclosure of nuclear content in eukaryotes implies that cells require sites in the NE that mediate exchange of material between nucleus and cytoplasm, given that RNA synthesis occurs in the nucleoplasm and protein synthesis occurs in the cytoplasm. The nuclear pore complexes (NPCs), 125 MDa macromolecular complexes, are responsible for this transport. [79, 80]. In vertebrates, NPCs consitst of 50-100 proteins, termed nucleoporins. Each NPC can be divided into three structural elements: cytoplasmic fibrils, the central core, and the nuclear basket. Transport through NPCs can be rapid (i.e. several hundred molecules/pore/second) by either passive diffusion of relatively small

molecules or active transport of complexes of up to several megadaltons. Each pore facilitates both import and export, using import and export receptors and small GTPases that travel with their cargoes through the pores. These cargoes include proteins and various types of RNA and RNP molecules [79, 80].

Dynamics of the nuclear envelope

Each cell division requires the complete breakdown and re-assembly of the NE, implying that the NE must have a high level of plasticity. In the early stages of mitosis, nuclear proteins are distributed throughout the cytoplasm [81, 82]. Shortly after the onset of anaphase the nuclear membranes assemble around chromosomes. Fusion proteins of green fluorescent protein (GFP) with the INM proteins LBR and emerin and the lamina protein lamin B1 demonstrated that these proteins localize to the daughter chromosomes in this very early telophase-stage [81, 82]. Despite its tremendous size and complicated molecular architecture, the NPC also breaks down and reassembles at cell division in vertebrates within a short period of time [79]. The timing of recruitment and spatial location of several nucleoporins coincides with recruitment of emerin and LBR to the daughter chromosomes. Nuclear import is recovered slightly later and chromosome decondensation occurs only after this recovery [82].

In interphase nuclei, kinetic studies of Lamin C and B1 fused to GFP show a very slow exchange rate when integrated into the NE, whereas nucleoplasmic GFP-lamin C diffuses rapidly. This has been interpreted to be due to the formation of higher-order polymers by lamin proteins and supports the view that the lamina is required for maintaining nuclear size and shape [81, 83]. Although incorporation and exchange of lamins is slow, the lamina itself shows considerable dynamic movements through folding and indentations. In addition, central core components of NPCs were found stably attached to a flexible lamina framework, but peripheral components can exchange rapidly with an intranuclear pool [84, 85].

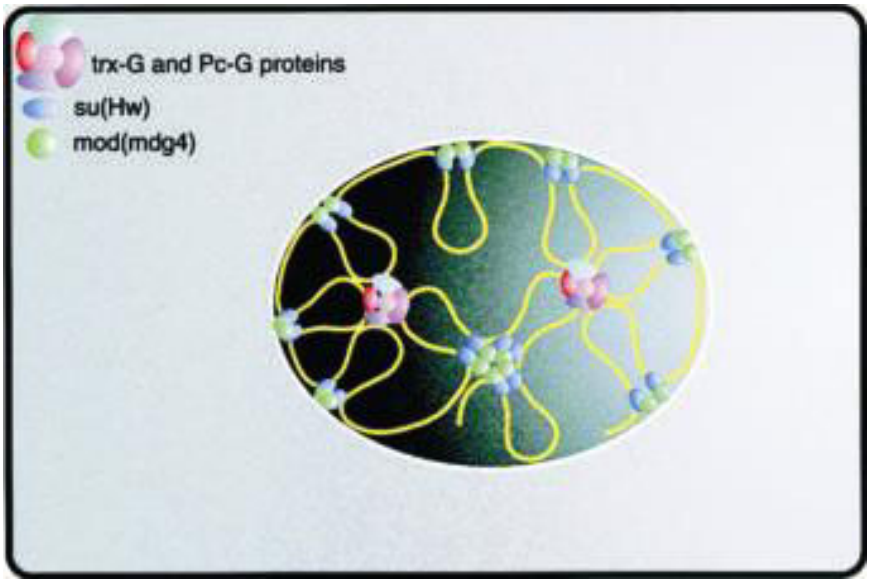


Figure 1.3.2. Schematic model that illustrates the role of *trx*G proteins in the function of the *gypsy* insulator higher-order chromatin domains.

The diagram represents a cell with a nucleus (oval). The chromatin fiber is represented as a yellow line, and proteins are represented as ovals: Mod(mdg4) (green), SuHw (dark blue) and various TrxG and/or PcG proteins (all other colored ovals). This figure was adapted from [92].

The nuclear envelope and transcription

Lamins bind to DNA, chromosomes and histones *in vitro* and the INM proteins LBR and LAP2 also bind to chromatin. Changes in lamin expression often correlate with transitions in development or differentiation, when gene expression patterns change. Thus, it has been proposed that the NE might influence the capacity for gene expression [78]. However, the actual effects of the NE on transcriptional regulation are largely unknown, while conflicting data have been published. Several indications suggest that the NE represents a repressive chromatin environment. First, heterochromatin is preferentially located near the envelope [78], while in the late stages of S phase, this heterochromatic DNA is released from the NE, moves to replication centers inside the nucleus and then returns to the envelope [86]. Second, it was shown that the yeast *HMR* locus, with a defective silencer element, exhibited improper silencing unless it was tethered to the nuclear periphery [87]. Third, native yeast telomeres, which have a heterochromatic structure, are highly dynamic but remain within a restricted volume adjacent to the NE during interphase [88-90]. Fourth, a nuclear periphery protein (Esc1) is a component of a pathway that functions to localize silencing complexes to the nuclear periphery [91].

In *Drosophila* a link between the NE and transcriptional activation, rather than silencing was observed. Here, a *cis*-regulatory DNA element, the *gypsy* transposon insulator, localizes to the NE in conjunction with *trans*-acting factors Su(hw) and mod(mdg4) to form so-called insulator bodies. A mutation in Su(hw) that destroys its insulator capacity causes a detachment form the NE. In addition, there was a correlation found between the organization imposed by the *gypsy* insulator and the ability of genes to be transcribed. Heat-shock response (i.e. heat-shock genes are turned on, the rest of the genes is turned off) causes redistribution of Su(hw) and mod(mdg4) in a diffuse pattern and reshuffling of insulator-linked DNA into the interior of the nucleus due to alterations in the chromatin organization imposed by the *gypsy* insulator [92]. Thus, *gypsy* may be required for establishing higher-order chromatin domains (figure 1.3.2), association with the NE and regulating gene expression.

In mammalian cells, the results are conflicting. In general, actively expressed genes are not found associated with the NE (reviewed in [93]). For instance, immunoglobulin (Ig) loci are selectively activated for transcription during B lymphocyte development. Ig heavy (H) and Igκ loci are preferentially positioned at the nuclear periphery in progenitors and pro T-cells but not associated with heterochromatin. In transcriptionally active pro-B cell nuclei they are centrally configured in the nucleus [94]. In addition, gene poor chromosomes are more likely to be at the nuclear periphery, while gene-rich chromosomes are localized to the nuclear centre [95, 96]. On the other hand, chromatin at the nuclear periphery is more DNase-sensitive than 'bulk' chromatin, probably reflecting transcriptionally active chromatin domains. The nuclear periphery has been proposed to function as a compartment for the spatial coupling of transcription and nucleocytoplasmic transport. Highly expressed genes might associate with NPCs to facilitate export of their mRNA [97-99]. Furthermore, it has been shown that a nuclear lamin mutant has defects in transcript localization (reviewed in [93]). Lamins are distributed throughout the nucleoplasm and disruption of the nuclear network formed by lamins is accompanied by a dramatic reduction in transcription by RNAPII that involves TBP, but does not detectably inhibit RNA polymerases I and III [100]. The authors suggest that lamins may act as a scaffold upon which the RNAP II basal transcription machinery is organized. Finally, a growing number of transcription factors either co-localize with the nuclear lamina or interact with proteins anchored to the lamina [101]. It should be noted in this context that expressed globin genes do not localize to the NE (Mariëtte van de Corput, personal comm.).

Thus, the NE may be involved in the establishment or maintenance of higher-order chromatin domains. On the other hand, the NE may be involved in a compartment dedicated to efficient transcription and subsequent nucleocytoplasmic transport, putatively located near NPCs. It is also

possible that the higher-order chromatin and nuclear compartment functions of the NE co-exist in a single nucleus.

1.3.2 Nuclear bodies

Nuclear bodies were the first substructures that were designated compartments. Their discovery depended mostly on studies involving light and electron microscopy. When light microscopy advanced into fluorescent and confocal microscopy, the number and types of nuclear bodies rapidly increased. Studies based on fluorescent fusion proteins, *in situ* hybridization of RNA and DNA, and immunolabeling of proteins, showed foci with an accumulated set of proteins (figure 1.3.3) and, in some cases, particular chromatin domains. Here, the best-characterized nuclear bodies are discussed.

The nucleolus

The nucleolus is assembled around clusters of tandemly repeated ribosomal RNA genes (rRNA genes) which are transcribed by RNAP I. The human rRNA genes are located in approximately 180 copies of tandemly arranged rDNA repeats on five separate chromosomes. These tandem arrays form the nucleolar-organizing regions and are the basis of the structural organization of the nucleolus, responsible for targeting of all processing and assembly components required for the nucleolus' major function, ribosome biosynthesis (reviewed in [102]). Morphologically it can be separated into three distinct substructures: fibrilar centres (FCs), dense fibrilar components (DFCs), and granular components. A typical human nucleolus is composed of ~30 FCs, each accommodating about four rRNA genes. They are surrounded by DFCs, while the granular components radiate out of the DFCs (figure 1.3.4). High-resolution *in situ* hybridization and Br-UTP-incorporation studies revealed that active rRNA genes are restricted to the periphery of the FCs, excluding their interior. The primary transcript (pre-rRNA) is approximately 13 000 nucleotides, known as 45S rRNA, and it is cleaved into 18S, 5.8S, and 28S rRNA [103]. Pre-rRNA processing is initiated in DFCs and finished in the granualar components [102, 104].

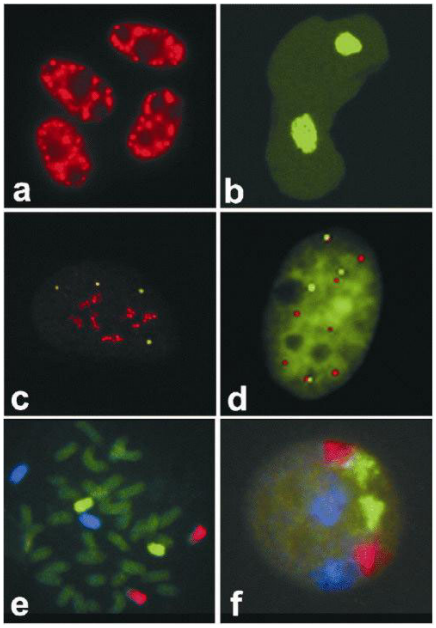


Figure 1.3.3. Nuclear compartments visualized by fluorescence microscopy.

Shown are splincing factor compartments (A, red), the nucleolus (B, green), the Cajal bodies and the nucleolus (C, green and red, respectively), PML bodies and Cajal bodies (D, red and green, repectively), and several mouse chromosomes (E) during metaphase. In interphase nuclei, chromosomes appear as finite and mutually exclusive entities, referred to as chromosome territories (F). Here, chromosomes 13 (blue), 14 (red) and 15 (green) are shown at both cell cycle phases (E and F). This figure was adapted from [102].

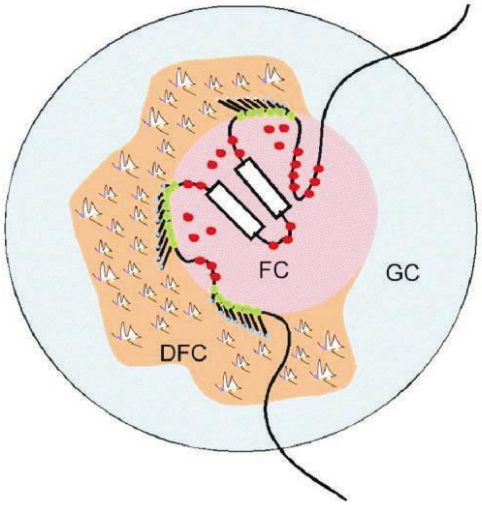


Figure 1.3.4. Organization of the nucleolar substructures.

The nucleolus is organized from chromatin-containing rRNA genes, which are anchored in the fibrillar centre (FC). Transcriptionally active rRNA genes are located at the periphery of FCs. Nascent pre-rRNA transcripts are entering DFCs, where they undergo a series of processing steps. Later processing and assembly steps occur in the granular component (GC). This figure was adapted from [102].

The rRNA genes are highly expressed, each active gene is associated with 100-120 engaged RNAP I molecules and transcripts are synthesized at a rate of ~2.5 kb per minute. Given that each transcript is about 13.3 kb, it takes approximately 5 minutes to transcribe an entire rRNA gene [105, 106]. Transcription of rRNA genes is essential for establishment and maintenance of the nucleolus. Upon inhibition of transcription the nucleolus disassembles (reviewed in [102]). Under these conditions only 11 different resident proteins were found by proteomics, whereas in transcribing nucleoli 271 different resident proteins were identified [107]. This demonstrates that transcription is the key to normal organization of the nucleolus. Furthermore, nucleolar components preferentially interact with each other and introduction of extrachromosomal rRNA genes triggers the spontaneous formation of novel nucleoli [102, 108]. Thus, interactions between rRNA genes and *trans*-acting factors appear to result in a transcriptionally active nucleolus.

Like the NE, mammalian nucleoli disassemble at the onset of mitosis and reassemble during telophase, indicating high structural plasticity. Partially processed pre-rRNA is preserved in association with processing components in the perichromosomal regions in particles called nucleolus-derived foci during mitosis. During telophase the nucleolus-derived foci disappear with a parallel appearance of material in the reforming nuclei. Prenucleolar bodies appear in nuclei in early telophase and gradually disappear as nucleoli are formed, strongly suggesting the transfer of prenucleolar body components to newly formed nucleoli. In interphase, nucleolar resident proteins are steadily and rapidly exchanged between the nucleoplasm and the nucleolus [77, 109]. The short residence times (order of seconds) and high exchange rate (thousands of molecules per second) demonstrate a large flux of proteins through the apparently stable nucleolus. RNAPI analysis shows that only 7 to 10% of the enzymes are actually engaged in transcription, leaving the vast majority moving freely at any given time [109]. The authors suggest a kinetic framework for RNAP I transcription *in vivo*. RNAP I machineries might be reassembled at each round of transcription in a sequential manner via meta-stable intermediates, each with increasing stability.

Transcription factories

Some research has focused on whether compartmentalization is involved in RNAP II and RNAP III transcription, analogous to the nucleolus. Several observations illustrate parallels between the different polymerases. First, in HeLa cells the number of nascent RNA transcripts per nucleus was determined to be ~90 000, with ~15 000, ~65 000, and ~10 000 being made by RNAPs I, II, and III,

respectively [105, 110]. Fluorescent detection of nascent transcripts in HeLa nuclei revealed a much lower number of foci of ongoing transcription, so-called transcription sites. It was estimated that each nucleus contains 5000 to 8000 RNAP II and about 2000 RNAP III transcription sites. Furthermore, many genes, but not all (e.g. globin genes [111]) were found associated with only one elongating complex at any given time [105, 110, 112]. The authors concluded from this that RNAP II and RNAP III transcription sites might represent a clustering of transcribed genes, the so-called 'transcription factories', that resembles rRNA gene clustering. Second, centers of RNAP II gene expression appear to be organized such that the chromatin template, transcription machinery and nascent product each occupy distinct zones (figure 1.3.5) [113] much like the structural organization of rRNA genes. Third, the kinetics of RNAP II enzymes are similar to that of RNAP I, the majority of nuclear RNAP II enzymes move freely throughout the nucleus and only 25% is engaged in transcription [114]. Fourth, several sequence-specific transcription factors fused to GFP were found to co-localize with transcription sites. They accumulated in bright foci that remained stationary in the nuclear space. However, kinetic studies showed a fast diffusion rate and only a fraction of the proteins was immobile [115]. Similar to the resident proteins of nucleoli, these transcription factors transiently accumulate at transcription sites, which results in visually stable structures. However, the analysis of co-localisation between transcription factors and transcription sites does not demonstrate whether transcription actually occurs once factors are recruited to a target gene [113].

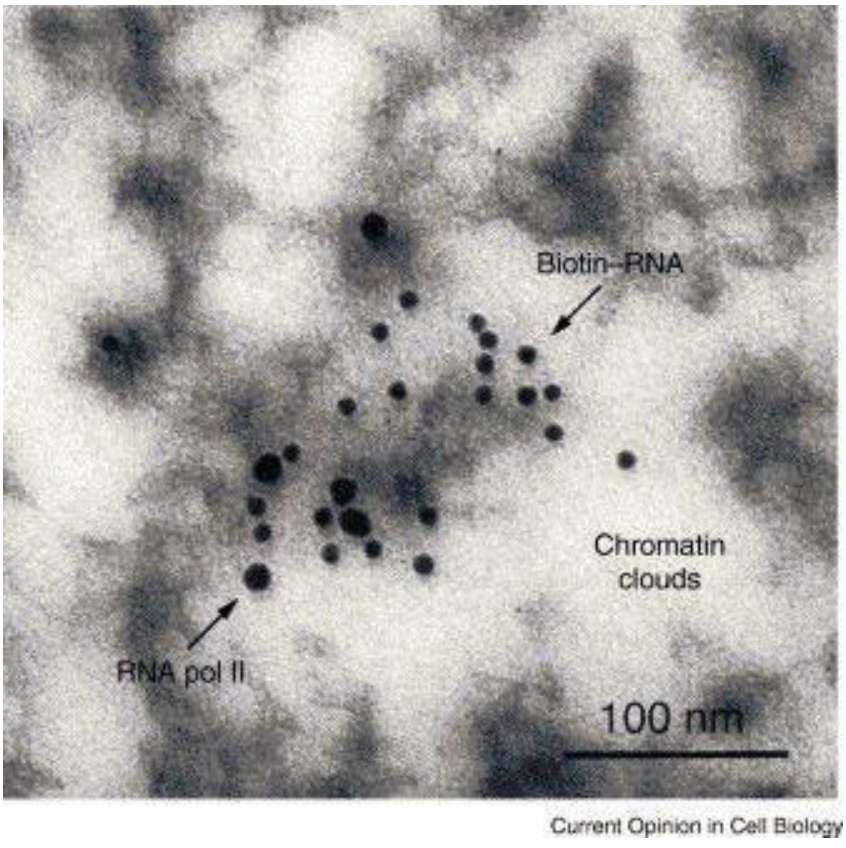


Figure 1.3.5. The anatomy of a transcription site.

Components involved in RNA synthesis are shown in a high-resolution electron microscopy (EM) image. In this example, the dense chromatin regions have a white/grey `cloud-like' appearance. Nascent transcripts are visualised with 5 nm gold particles. The transcription machinery is visualised with 10 nm gold particles. This figure was adapted from [113].

Splicing-factor compartments, Cajal bodies and PML bodies

In addition to the nucleolus, other morphological structures have been identified (reviewed in [102, 108, 116]). Three of these are described briefly: splicing-factor compartments (SFCs), Cajal bodies (CBs) and promyelocytic leukaemia oncoprotein (PML) bodies. SFCs or nuclear speckles are composed of dense granules (diameter 20-25 nm) connected by fibers. Its resident proteins are mainly pre-mRNA splicing factors. CBs are spherical structures (diameter 0.1-1.0 μm) that contain a large number of components, including spliceosomal snRNPs, and GTFs. Their function is unknown but transport/maturation of snRNPs and assembly site for the basal transcription machinery have been proposed. PML bodies are small spherical domains (diameter 0.2-1.0 μm) and functions have been implicated in terminal differentiation, transcription regulation, nuclear storage, growth control, and apoptosis. In addition to the PML protein, these bodies contain many other proteins.

In relation to transcription, CBs associate preferentially with specific genomic loci (snRNA genes and histone gene clusters). SFCs are not primary sites of pre-mRNA splicing, rather they provide storage and assembly of the splicing machinery. However, most active genes are found at the periphery of SFCs, which suggests that all RNA-processing machines are linked physically to the transcription process. Similar to what is observed in nucleoli and 'transcription factories', SFCs, CBs and PML bodies behave in a dynamic way. SCFs undergo continuous changes in shape, suggesting high internal dynamics [117]. CBs and PML bodies move in an ATP-dependent manner [118, 119]. A rapid and continuous flux of proteins between nuclear bodies and the nucleoplasm has also been demonstrated [118, 120].

1.3.3 Chromatin compartments

The interior of the nucleus is clearly not homogeneous, when examined with either light or electron microscopy. Two types of chromatin can be distinguished: one that remains condensed throughout the cell cycle and another chromatin fraction, which decondenses as the cell progresses from metaphase to interphase, referred to as heterochromatin and euchromatin, respectively (reviewed in [121]). These two types of chromatin can be considered nuclear compartments, each with its distinct properties (table 1.3.1) and resident proteins.

Euchromatin

The differences between euchromatin and heterochromatin are detectable at the DNA sequence level. Characteristically, euchromatin is largely composed of unique sequences that contain a relatively high to variable density of genes. In addition, it is relatively GC-rich mainly because CpG islands are found at the 5' end of many mammalian genes (approx 60%), including most house keeping genes and 40% of genes that have a tissue-restricted pattern of expression [122]. In CpG islands the dinucleotide CpG occurs with a 10-fold higher frequency than in the rest of the genome. In addition, the cytosine residues of CpG islands are unmethylated, while they can be methylated outside CpG islands ([123] and references therein). DNA methylation is associated with heterochromatin and transcriptional repression. Furthermore, CpG islands contain very little histone H1 (a histone involved in chromatin condensation) [124] and they are initiation sites for both transcription and replication of DNA [125]. Euchromatic domains have irregular nucleosomal arrays due to the presence of DNase I HSs. Euchromatic domains usually contain hyperacetylated histones and methylation of histones can be found in the form of H3-meK4 (see also figure 1.2.4) [124, 126]. In conclusion, euchromatin forms a decondensed or open chromatin compartment with many biochemical 'marks' associated with active transcription and high gene density.

Table 1.3.1 Distinctions between euchromatic and heterochromatic compartments.

Feature	Euchromatin	Heterochromatin
Staining/packaging in interphase	Dispersed/decondensed	Appears condensed, heteropycnotic
DNA sequence	Predominantly unique	Predominantly repetitive (satellites, derivatives of viruses, transposons, etc.)
Presence of genes	High/variable density	Low density
Meiotic (reciprocal) recombination	Normal frequency	Low frequency
Replication timing	Throughout S phase	Late S phase
Chromatin structure	DNase I HSs, irregular nucleosomes, accessible to nucleases	Loss of DNase I HSs, regular nucleosomal array, less accessible to nucleases
Activity state		
Euchromatic genes	Genes inducible	Genes silenced (variegated)
Heterochromatic genes	Genes silenced (variegated)	Genes inducible
Characteristic modifications	Histone hyperacetylation Histone H3-meK4 present Cytosine (CpG) hypomethylation	Histone hypomethylation Histone H3-meK9 present Cytosine (CpG) hypermethylation

Adapted from [126].

Heterochromatin

Heterochromatin is a more condensed form of chromatin and can be classified into two populations: constitutive and facultative heterochromatin. The former remains visibly condensed during interphase and is mainly located at pericentromeric chromosomal domains in mammals. Facultative heterochromatin is conceptually defined as developmentally regulated heterochromatin capable of undergoing a transition between the heterochromatic and euchromatic states [121]. This transition can occur during development and in differentiation pathways when a particular subset of the active genes is silenced. It can also occur in PEV when normally active chromatin can be silenced when placed next to heterochromatin. It is presently not clear whether facultative heterochromatin shares all of the structural features of constitutive heterochromatin.

Constitutive heterochromatin contains many repetitive sequences, particularly at the centromeres with long stretches of satellite repeats. Presence of long tandem arrays of a sequence motif will have the effect of generating relatively uniform sequence content over a large region. Recent data obtained in yeast indicate these may form heterochromatin as a result of an RNAi-type repression [127-129]. Three well-characterized covalent modifications that ‘mark’ heterochromatin have been described: hypoacetylation of histone lysine residues, histone methylation in the form H3-meK9 (see also figure 1.2.4), and DNA methylation at the cytosine residues. These modifications are associated with transcriptional repression and may be involved in packaging heterochromatin and stable silencing of euchromatic genes [53, 126]. Large-scale heterochromatinization of the genome is observed in terminally differentiated cells, such as plasma cells, and glial cells. In addition, the inactive X chromosome in mammals forms a densely stained structure (the Barr body) in which the core histones are hypoacetylated and H3 K9 is methylated [121]. In addition to transcriptional silencing, heterochromatin may also be required for chromosome stability during cell divisions. In *S. pombe* the heterochromatic protein Swi6 is required for efficient cohesion between sister centromeres [130]. Mice with a null mutation of *Suv39h*, encoding an H3-meK9 specific HMTase, no longer bind heterochromatin protein 1 (HP1) at the centromere and show increased chromosome instability with a higher frequency of non-homologous pairing in male meiosis [131].

Chromatin dynamics

In situ data demonstrated that repositioning of genes, proteins and *cis*-regulatory elements relative to a particular chromatin compartment occurs in response to alterations of transcriptional activity and mitogenic activity. The insertion of heterochromatin into the euchromatic *brown* gene (*Drosophila*) results in PEV and aberrant association of the gene and its homologous copy with the heterochromatin compartment in interphase nuclei [132]. In the nucleus of B lymphocytes Ikaros proteins (required for normal T, B, and NK cell development) localize to discrete centromeric heterochromatin containing foci and associate with transcriptionally inactive but not transcriptionally active genes. In addition, in developing T cells two gene loci were shown to relocate to centromeric domains following heritable gene silencing. These findings suggest that repressed genes are selectively recruited into centromeric domains [133, 134]. However, several transcriptionally repressed genes were not associated with centromeric heterochromatin in mature resting B lymphocytes. Mitogenic activation of these cells resulted in association with centromeric heterochromatin. This indicates that the spatial organization of genes in cycling and non-cycling lymphocytes may be different and suggests that locus repositioning to the heterochromatin compartment may be a feature of heritable gene silencing [134].

The dynamics of the chromatin template were monitored in living cells after the development of a Lac^O/LacR tagging tool. This tool originally consisted of 256 direct repeats of the bacterial Lac^O sequence that were introduced into the genome of cell lines as large tandem arrays due to gene amplification using dihydrofolate reductase (DHFR) as a selectable marker. Next, the Lac^O array was tagged with a fusion between GFP and LacR, which is the cognate DNA binding protein of Lac^O. This allowed a live analysis of these cells using fluorescent microscopy [135]. In yeast, this revealed movements of ~0.5 μm in ~10-second intervals for the internal chromosomal loci in G_1 -phase nuclei. Chromatin movement seemed to be undirected, sensitive to ATP depletion and to changes in metabolic status in the cell. Therefore, the authors proposed that the movements reflect the action of large ATP-dependent enzymes involved in transcription or chromatin remodeling [90, 136]. The same fast short-range movement has been reported for mammalian and *Drosophila* spermatocyte nuclei [137, 138]. However, it should be noted that this type of movement in mammalian nuclei restricts a locus to only ~$1/1000^{th}$ of the nuclear volume [137], while in yeast half the radius of the nucleus can be crossed with the same type of movement (figure 1.3.6) [90, 136]. In addition, loci at nucleoli or the nuclear periphery are significantly less mobile than other, more nucleoplasmic loci. Although the difference in linear distance is merely 1.5 to 2-fold, this corresponds to a further restriction of 4 to 8-fold in the nuclear space [137]. Thus, chromatin motion appears to be confined to a limited volume of the nucleus and the rapid random walk movements are thought to reflect constrained diffusion. In *Drosophila*, an additional slower component was detected that is limited to a much larger chromosome-sized domain [138]. This is in agreement with fluorescence *in situ* hybridization analyses that have demonstrated that chromosomes occupy a finite and mutually exclusive fraction of the nuclear volume, the so-called chromosome territories (reviewed in [99]).

The effects of transcriptional activation on chromatin were also examined in living cells with the Lac^O/LacR tagging tool. A 90 Mbp array of Lac^O sequences stained with LacR-GFP as condensed foci of approximately 1 μm in diameter [135], whereas a fusion of LacR-GFP to the potent VP16 transcriptional activation domain resulted in a dramatic decondensation of these fluorescent foci [139]. A similar dramatic decondensation upon transcriptional activation was reported elsewere [140]. Decondensation was accompanied by an increase in acetylation of all four core histones, recruitment of several HATs, SWI/SNF components, and activation of transcription. Interestingly, inhibition of transcription with α-amanitin does not block the decondensation [139, 141].

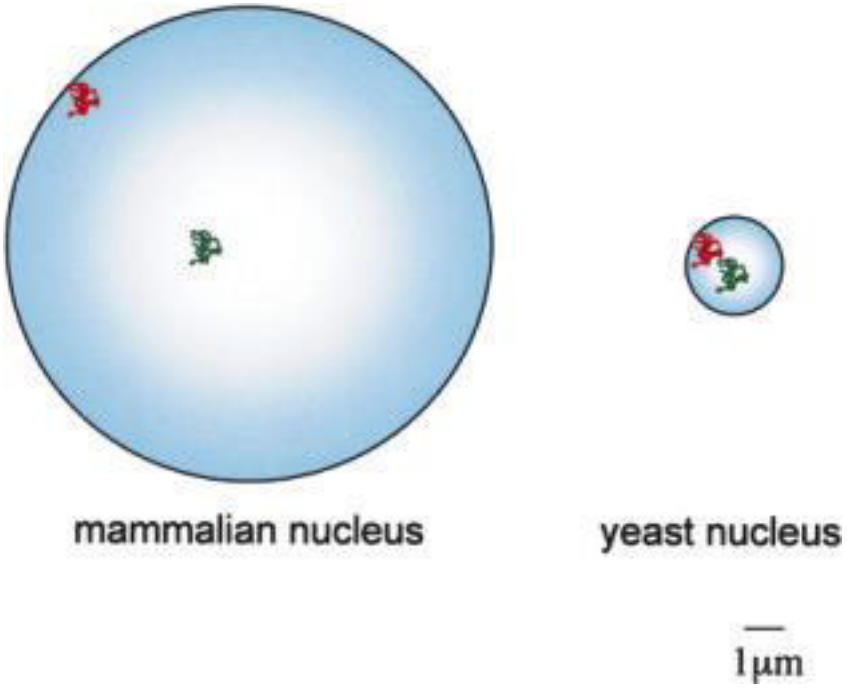


Figure 1.3.6. Constrained chromatin motion in mammalian and yeast nuclei.

The random-walk, but constrained (0.5 μm average range) chromatin motion of a locus in both mammalian and budding yeast nuclei allows it to sample only a portion of the nuclear volume. A locus positioned close to the nuclear periphery (red) may encounter a local environment that contains different *trans*-acting factors than a locus distant from the periphery (green). For instance, the nuclear periphery is associated with gene silencing and presumably a local accumulation of silencing protein (see paragraph 1.3.1). A centrally located locus is less likely to encounter these 'silencing' regions with its limited mobility. However, it should be noted that data regarding the transcriptional activity of the mammalian nuclear periphery are conflicting (paragraph 1.3.1). This figure was adapted from [273].

Dynamics of chromatin proteins and interacting proteins

The most prominent proteins of chromatin are the core histones (H2A, H2B, H3, and H4 see figure 1.2.1) that form an octamer with the DNA wrapped around this core. They are largely responsible for the repressive ground state of chromatin [1] and their dynamic behavior has recently become of great interest. Kinetic experiments revealed that the inner core of the nucleosome (containing H3 and H4) is very stable, whereas H2B on the surface of active nucleosomes exchanges both rapidly and slowly. The rapidly exchanging fraction was proposed to represent H2B in transcriptionally active chromatin [142].

In addition to the core histones, histone H1, HMG proteins and HP1 are also chromatin proteins, which are responsible for higher-order chromatin structure. Almost the entire population of histone H1 is bound to chromatin at any one time, however, it is exchanged continuously between chromatin regions. The behavior of histone H1 can be summarized in a 'stop-and-go' model. A histone H1 molecule binds chromatin for about 1 to 2 minutes, falls off, and then diffuses freely through the nucleoplasm for a short period of time until it encounters a free binding site [143]. HMG proteins also behave in a 'stop-and-go' fashion although their exchange rate is faster (in the order of seconds) and they have a significantly larger unbound fraction that moves rapidly throughout the entire nucleus [77]. Recent studies on the binding of HP1 protein in heterochromatin indicate that the individual

components within these domains are not bound stably, but in constant flux. The residence time of HP1 in euchromatin was 5 to 90 seconds and in heterochromatin between 60 and 200 seconds [144, 145]. The residence times of histone H1 and HP1 are further reduced by alterations in the histone hyperacetylation and in cells lacking the histone H3K9 methylase Suv39 [143, 144].

Transcription factors directly interact with chromatin, since they bind specific DNA sequences. Hence, their binding kinetics to specific DNA sequences may provide insight in how they activate or repress transcription. With the use of a tandem array of reporter elements and a glucocorticoid receptor labeled with GFP, targeting of the receptor to response elements in live mouse cells was observed. Photobleaching experiments provide direct evidence that the glucocorticoid receptor-GFP as well as its co-factor (GRIP-1) fused to GFP undergo rapid exchange between chromatin and the nucleoplasmic compartment [146, 147]. Similar dynamics have been observed for the intranuclear estrogen receptor (ER) and the steroid receptor co-activator 1 (SRC-1) [148]. Similarly the RUNX proteins and the AhR/ARNT transcription factor complex were shown to behave in a dynamic manner similar to that observed for the glucocorticoid and estrogen receptors [115, 149]. Thus, transcription factors appear to move very rapidly throughout the entire nucleus and their interaction with target sites in chromatin is transient. Considering the high DNA content and the large amounts of RNAs and proteins, one might intuitively expect the nucleus to be a viscous environment, however photobleaching experiments have now shown that many proteins are highly mobile within the nucleus [150]. The dynamic properties of nuclear proteins also suggest that a stochastic mechanism provides the basis of gene expression and nuclear architecture [150]. In such a dynamic environment, the fate of a chromatic domain is likely determined by a competition for binding between activators and repressors. Binding is dependent on sequence-specific elements recruiting these proteins, the residence time of each factor at its cognate binding site in chromatin, and the relative local concentrations of both activator and repressor [151].

The nuclear organization depends on dynamic interactions between components with constrained and high mobility. The rapid exchange of proteins binding to or dissociating from macromolecules and moving in or out of bodies generates a dynamic but stable nuclear organization of the larger structures. A striking feature of many proteins found in nuclear compartments is the presence of self-interacting domains (reviewed in [152]). These characteristics are supportive of a concept of self-organization in the formation of nuclear structures. In a self-organizing system, the interactions of its molecular parts determine its architectural and functional features. This may provide an elegant mechanism not only to effectively concentrate factors where they are needed to drive biochemical reactions, but also to segregate factors away from sites where they are not required [152]. This brings up the question whether *cis*-regulatory DNA elements contribute to the nuclear organization because these elements bind a multitude of factors. In the nucleolus, active rRNA genes and presumably their *cis*-regulatory DNA elements are restricted to the periphery of the FCs and CBs associate preferentially with specific genomic loci. The AhR/ARNT complex binds to specific *cis*-regulatory DNA elements of numerous target genes and accumulates in multiple bright foci that co-localize with active transcription sites. This suggests that gene sequences might provide nucleation sites for the formation of nuclear structures [152].

1.4 Globins

The globin genes have a long history as a model system to study gene regulation. One compelling reason is that mutations in or deletions of globin genes are responsible for the most widespread genetic disorders, thalassemia and sickle cell anemia [153]. The products of the globin genes, globin mRNA and hemoglobin protein (see box: hemoglobin) occur in large amounts in erythroid cells and can be isolated easily. This contributed to the fact that globin genes and disease were among the first to be studied by biochemical and molecular biological methods. In the context of transcriptional regulation, the globin genes are an important system to study tissue-specific and developmentally-regulated transcription. Over the past, the globin genes have played a major role in identifying *cis*-regulatory DNA elements that contribute to the proper regulation of transcription.

Box: hemoglobin
Hemoglobin is a heterotetramer that contains two α-globin peptide chains and two β-globin peptide chains. Each globin chain binds one heme-group that binds oxygen in a reversible manner. Hemoglobin is abundant in erythrocytes that circulate in the blood stream and functions as a carrier of oxygen from the lungs to respiring tissues. Conversely, it can transport carbondioxide from the respiring tissues to the lungs.

1.4.1 The human and mouse β-globin loci

The globin genes encode peptide chains that form a heterotetramer called hemoglobin. In vertebrates, two types of globin genes exists due to gene duplications [154], the α-like globin genes and the β-like globin genes. This introduction is focussed on the β-like globin genes, since the research data in this thesis have been obtained from the human and mouse β-globin loci (see chapters 2, 3, and 4).

The β-like globin genes encode relatively small proteins of 146 amino acids. All genes have three coding regions (exons) separated by two intervening sequences (introns) and are immediately flanked by regulatory elements *in cis*. These elements are sufficient and necessary for tissue-specific expression and developmental timing of the individual genes. They include promoters, enhancers and silencers; the latter are counterparts of enhancers and enable the silencing of genes linked *in cis* in a distance and orientation independent manner. Most of the data investigating there functional activities are based on the human globin genes often inserted as transgenes integrated in the mouse genome.

The globin promoters can be detected as DNase I HSs in the chromatin of erythroid cells. The predominant sites in definitive erythroid cells are located at the 5' sides of the globin genes within 200 bp of the respective transcriptional start site. In human fetal livers, the HSs of the fetal and adult genes can be identified, while in bone marrow only the sites near the adult genes are present [8]. Thus, the DNase I hypersensitive state of globin promoters is controlled developmentally. All globin promoters contain similar but not identical TATA, CAAT and CAC box sequences (figure 1.4.1). The LCR strongly stimulates expression from a promoter possessing only a TATA box, however, the combination of the LCR and the CAC/CAAT elements is necessary for erythroid-specific transcription [155]. Like the β-globin promoters, the enhancers are also detectable as DNase I HSs [8]. They are located in close proximity or even within the β-globin like genes, 3' of the $^{A}\gamma$ and β-globin genes as well as in the second intron of the β-gene [13, 14, 156]. Deletion of the 3' β-globin enhancer showed decreased transcription of the β-globin gene [157]. Silencer elements are present at various positions in the β-globin locus, 5' of ε and the γ genes [158, 159].

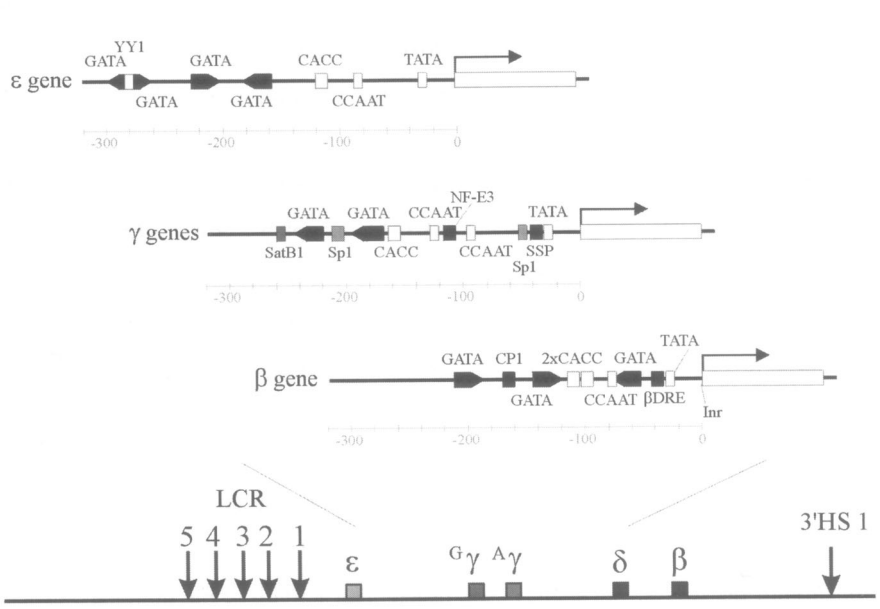


Figure 1.4.1. The promoter regions of the human ε-, γ-, and β-globin genes.

Shown are the locations of conserved *trans*-acting factor binding sites. The type of *trans*-acting factor that associates with a particular type of binding site can differ per promoter.

Genomic organization

Humans and mice have a similar, but not identical genomic organization of their β-globin loci (figure 1.4.2). The *cis*-regulatory DNA elements and genes are highly conserved between the two loci [160-162]. Both have approximately a 40% GC-content and almost an identical percentage of repetitive DNA (~30%) [162]. The human locus contains , they are arranged starting at the 5' side with ε, ^{G}γ, ^{A}γ, δ, and β to the 3' side. In the mouse, a similar arrangement can be found of its four genes, 5'- εy-βh1-β^{maj}-β^{min}-3'. Regions of synteny containing olfactory receptor (OR) genes are flanking the locus [161, 162]. Both loci have erythroid-specific distally located *cis*-regulatory DNA elements, including an LCR, 3'HS1 [163], and upstream located HSs. In mouse, these sites are located 85/84 kb (5' HS-84/-85) and 62/60 kb (5' HS-60/-62) upstream of εy [39, 164]. In humans, a homolog of the mouse 5' HS-62 was found at 110kb and an additional site at 107 kb upstream of ε (5' HS-107/-110) [39].

The first indications of the existence of a distally located *cis*-regulatory DNA element came from a Dutch family carrying a thalassemia deletion that did not express any β-like globin genes from the affected chromosome. In this deletion, the γ and δ genes were absent but the β-globin gene was still present toghether with its proximal *cis*-regulatory DNA elements (figure 1.4.2). The unaffected chromosome did express globin genes, hence it was concluded that a deletion far from the β-globin gene resulted in the suppression of its activity rather than a defect in *trans*-acting environment [22, 165]. It later transpired that the deletion included a series of DNase I HSs located upstream of the globin genes [163, 166]. In transgenic mice, linkage of these sites to a β-globin gene resulted in tissue-specific, position-independent and copy-number dependent expression [16], while this was not the case for globin transgenes without an LCR [7, 13, 14]. The mouse LCR was identified on basis of sequence homology between the human and mouse loci [167]. An analysis of the human LCR chromatin in transfected mouse erythroleukemia (MEL) cells suggested that the formation of the HSs

precedes β-globin transcription [168]. However, the LCR needs to be linked to an active promoter to stay hypersensitive [169-171]. HS formation requires binding sites for multiple transcription factors, including GATA-1, NF-E2, EKLF, and Sp1 [172-174]. Each HS holds an array of binding sites for transcription factors with similarities among the individual sites (figure 1.4.3) [172, 175, 176] although their functional properties are not equivalent.

In humans, 5' HS1 directly to a transgene does not result in expression [177], but contributes in context of the complete locus [28]. 5' HS2 shows classical enhancer properties [178] and contributes equally to overall expression of the locus throughout development [177]. The enhancer property is mediated by a tandem NF-E2 binding site that is not found as a tandem array in any of the other HSs [179]. Only multi-copy integration of 5' HS2 linked transgenes is sufficient to drive position-independent and copy-number dependent expression [15]. HS3 is the most active site during the embryonic period, and the only site capable of high-level expression of the gamma genes during fetal hematopoiesis [177]. It can control transcription of single copy integrations and establish/maintain DNase I HSs, which suggests a chromatin opening or remodeling activity [180]. 5' HS3 activity largely depends on EKLF [173, 181]. The highest activity of 5' HS4 is observed during the adult stage of development [177]. Insertion of the 5' HS4 element in place of 5' HS3 results in significant expression changes at every developmental stage [182]. 5' HS5 was originally identified as a constitutive HS with boundary/insulator activities [163, 183]. 5' HS5 also contains binding sequences for an enhancer-blocking transcription factor, called CTCF [184]. However, recently it was shown that 5' HS5 is erythroid cell-specific with enhancer-blocking activities only at the embryonic stages of development but not at later stages [35]. Thus, the LCR is composed of multiple HSs with separable functions. A possible *in vivo* mechanism for LCR activity was proposed based on the chromatin opening activity of 5' HS3. In this model, chromatin at the LCR is first opened by *trans*-acting factor activities at 5' HS3 and subsequently spreads throughout the LCR. Protein-protein interactions between factors bound at the individual sites were postulated to establish a 'holocomplex' [180]. This model is supported by the observation that an incomplete LCR often shows reduced levels of expression in transgenic mice and loses its properties of position-independent and copy-number dependent expression when integrated into heterochromatin [28]. In addition, 5' HS3 activity largely depends on EKLF, which interacts *in vitro* with the chromatin remodeling factor E-RC1 [63, 173].

In mice, the different properties of the individual HSs are less clear. Targeted deletions of 5' HS1, 5' HS2 and 5' HS4 reduce expression levels of all the genes similar to what was observed for the human sites [185, 186]. A targeted deletion of 5' HS3 results in a minimal reduction of embryonic genes and 30% reduction of adult genes [187], a deletion of 5' HS5 has minimal effects on expression [188]. In all cases, the deletion did not affect the formation of the remaining sites [188, 189]. This suggests that in the endogenous mouse locus no dominant or initiating site whose formation must precede the other HSs exists. The discrepancies found between human and mouse may reflect differences in the functional assays used, transgenic mice versus targeted deletions or the fact that the sites show some difference in numbers and affinities of the individual factor binding sites. Alternatively, the endogenous mouse locus may contain redundant elements that are missing from the transgenic constructs. Recently, two more 5' HSs of the LCR have been identified, 5' HS6 and 5' HS7 [161]. Targeted deletion of 5' HS6 from the endogenous mouse locus has minimal effects on expression [188] with again no effect on formation of other sites [189]. The functional activities of equivalent human elements have not yet been determined. Lastly, the functional significance of the distally located HSs (see figure 1.4.2) was until recently unknown. Novel data imply an active role of some of those elements in β-globin gene transcription (chapter 3 and 4) [190, 191].

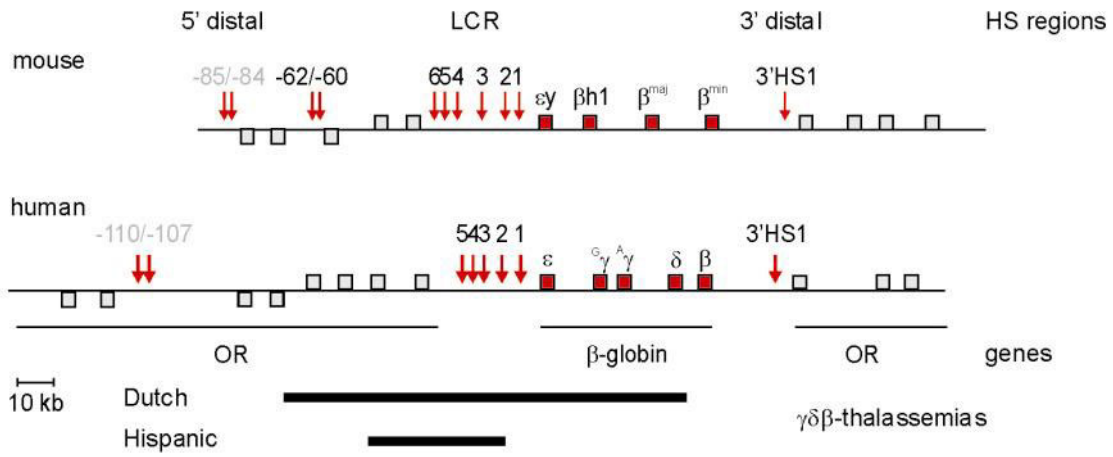


Figure 1.4.2. The mouse and human β-globin loci.

The β-globin loci of man and mouse are shown schematically. Boxes above the horizontal lines represent genes transcribing from left to right, boxes below represent genes transcribed in the opposite direction. The globin genes are indicated (red) and corresponding erythroid cell-specific DNase I HSs are represented by red arrows. The *cis*-regulatory DNA elements of the loci are positioned in distinct regions. Each region contains one or more HSs and names of each site are specified. HSs that are involved in globin gene expression are labeled in black and HSs with currently unknown function in grey. Olfactory receptor genes (OR, grey) flank and partially overlap both loci. Below the human locus the size of two deletions (Dutch and Hispanic) that are known to cause γδβ-thalassemia are indicated (see text for details).

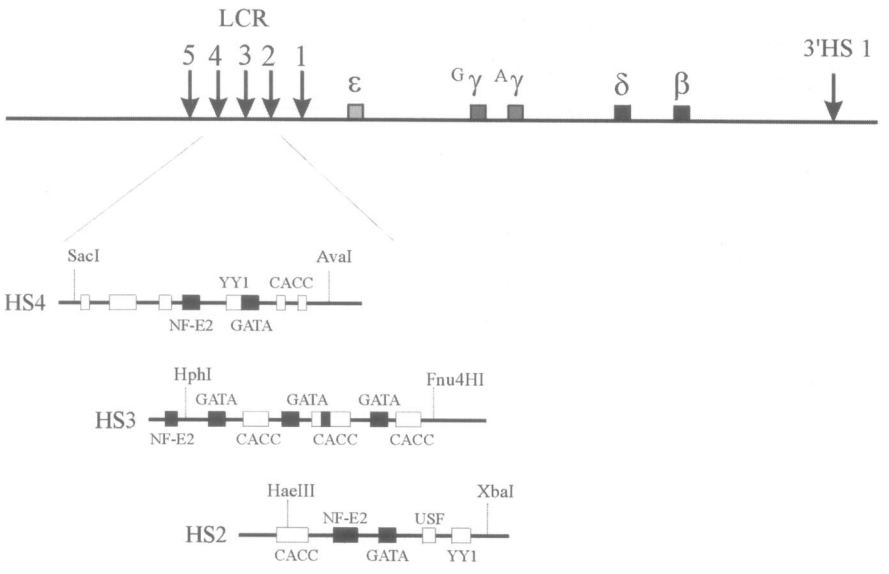


Figure 1.4.3. The core regions of 5' HS2 , 3, and 4 of the human LCR.

The binding motifs of various sequence-specific transcription factors are depicted.

Chromatin state

Recently, it was reported that a domain of DNase I sensitivity in the endogenous mouse locus extends from approximately 10 kb upstream of 5' HS-60/-62 to a few kb downstream of 3' HS1 [39]. The human locus is also more sensitive to DNaseI I than 'bulk' DNA in erythroid cells [2]. In a Hispanic thalassemia patient, a deletion was found that extended from 9.5 kb to 39 kb upstream of the ε-globin gene, leaving the distally located HSs (i.e. 5' HS-107/-110 and 3' HS1), 5' HS1 of the LCR and all the genes intact (figure 1.4.2). However, the genes were found to be inactive [23]. In addition, the remaining parts of the locus had become DNase I-resistant and HSs were not formed [192]. These findings support the transgenic data that the LCR is required for chromatin opening of the whole locus. In contrast, deleting the human endogenous LCR in an erythroid background did not show any changes in chromatin structure of the human locus [25]. The authors interpreted this as the LCR being necessary to open, but not maintain an open chromatin formation. Interestingly, a targeted deletion of the mouse LCR shows that the locus is still DNase I sensitive and that the remaining HSs were present. This was even detected after germline passage of the deletion. This suggests that elements elsewhere in the mouse locus are sufficient to establish and maintain an open chromatin conformation [26]. The transgenic experiments have clearly demonstrated that the human LCR mediates the open chromatin structure of the β-globin locus and is supported by the data on patients. However, these are in apparent contradiction with the data on the endogenous mouse locus and the deletion of the human LCR after introduction in erythroid cells. It is presently not clear why, although recent data on erythroid precursor cells show that there are intrinsic differences in chromatin modifications between the two loci. Remarkably, this difference is maintained in the human locus in transgenic mice suggesting that the primary DNA sequence rather than the organism determines the difference in epigenetic code [193].

DNA methylation at the cytosine residues is different in nonexpressing versus expressing erythroid cells. In expressing cells, the regions surrounding and including the genes showed little modification, which was not the case in neighboring DNA and the entire DNA locus in nonexpressing cells [194]. The mouse β-globin locus showed erythroid cell-specific hyperacetylation and histone H3 MeK4 at its *cis*-regulatory DNA elements. In definitve erythroid cells, four regions within the locus exhibited hyperacetylation: 5' HS-62/-20, the LCR and transcriptionally active β^{maj} and β^{min}-globin genes [39, 195]. Other regions, including the inactive genes and 3' HS1, were modestly enriched in acetylation and methylation, while the region encompassing the OR genes in between LCR and 5' HS-62/-60 showed similar levels as inactive control genes [39]. In embryonic yolk sac, which expresses εy and βH1, the LCR and both active and inactive promoters were hyperacetylated [195]. This is analogous to the developmental changes in DNase I hypersensitivity found at the promoters of the human β-globin genes [8]. Chromatin analyses of erythroid precursor cells, which express the globin genes at basal levels [196, 197], showed that the β^{maj} promoter have low to moderate acetylation levels and the mouse 5' HS3 and β^{maj} promoter are already accessible to DNase I. In fully differentiated cells, which express the genes at high levels, acetylation and accessibility were further increased [193]. In the human locus, acetylation levels were tested in MEL cell hybrids. Human chromosomes carrying either a wild-type locus or a locus with a deleted LCR showed a similar acetylation levels throughout the locus. Conversely, a chromosome carrying the Hispanic deletion was hypoacetylated [198]. The chromatin of the LCR in precursor cells is acetylated, with the exception of 5' HS3, and accessible to DNase I. Accessibility and acetylation, including 5' HS3, further increased in mature erythroid cells. Chromatin at the human γ- and β-promoters revealed a curiously high level of histone H3 acetylation in precursors and this decreased during differentiation. Conversely, the β promoter is inaccessible to DNase I in precursors, while upon differentiation accessibility increased [193]. The authors proposed that this transient chromatin acetylation may be involved in gene-specific

potentiation in precursor cells (i.e. before extensive chromatin remodeling and transcription take place in erythroid cells).

1.4.2 Hemoglobin gene switching

The β-globin genes have a well-characterized developmental transcription pattern. In all species that contain β-globin genes, the switch in globin gene expression coincides with changes in morphology of the erythroid cell, the site of erythropoiesis, and hemoglobin composition [19]. Hemoglobin gene switching happens differently between humans and mice (figure 1.4.4). In humans, a switch from embryonic (ε) to fetal (γ) globin gene expression and a second one from fetal (γ) to adult (δ/β) globin expression can be observed, whereas in mice expression only switches from embryonic (εy/βh1) to adult (βmajor/βminor). The order of globin genes reflects their expression pattern during ontogeny (figure 1.4.2) [19, 199]. The expression of the embryonic genes occurs in primitive cells, which are derived from the embryonic yolk sac. In contrast the fetal/adult genes are expressed in definitive cells, which are in origin derived from stem cells born in the AGM region of the developing embryo [200].

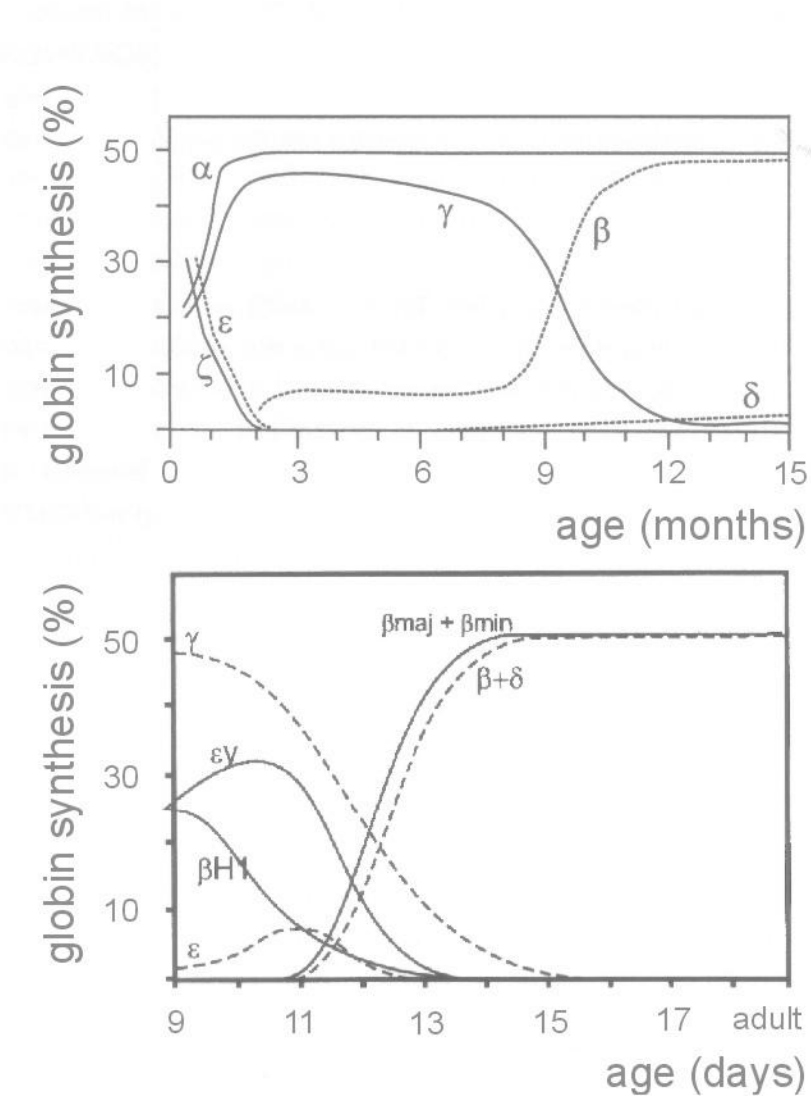


Figure 1.4.4. Developmental regulation of globin synthesis.

Top panel shows the expression levels of the individual human globin genes during development. The α- and ζ-globin genes are located in the α-globin locus. Bottom panel shows the expression levels of the individual mouse globin genes (solid lines) and human transgenes (dotted lines).

Molecular mechanism of switching

Several studies have led to the proposal of a dual molecular mechanism of hemoglobin gene switching: autonomous gene control and gene competition for direct interaction with the LCR [201, 202]. The concept of autonomous gene control is largely based on human globin genes in transgenic mice. Early experiments showed that individual human transgenes with just proximal *cis*-regulatory DNA elements (i.e. promoters, enhancers and silencers) express tissue-specifically and with the correct developmental timing, albeit in a position dependent manner [7, 14, 21, 203]. Transgenes of the ε and γ-globin genes, including their proximal *cis*-regulatory DNA elements, linked to an LCR are properly silenced at later stages of development [204, 205], silencing elements reside in their promoter regions. The 5' ε-element reduces expression of CAT-reporter construct [158], deletion or mutations lead to improper developmental timing of the ε-globin gene [206, 207]. In addition, this region is necessary for ε and γ activation at the embryonic stage [157]. Suppression of ε transcription requires a direct repeat element in its promoter that can be bound by a novel protein complex *in vitro* [208, 209]. In humans, point mutations in the promoters of the γ genes give rise to non-deletion hereditary persistence of fetal hemoglobin (HPFH), which is a genetically inherited condition that results in continued expression of the fetal γ-globin genes in adult life [19]. A targeted deletion of the endogenous mouse LCR gave reduced expression levels, but tissue-specificity and developmental timing is not affected [26]. Altogether these data demonstrate that the human ε- and γ-globin genes are accurately silenced independent of LCR functioning. The proximal regions of the embryonic/fetal genes are sufficient and necessary for autonomous silencing.

The first indications for gene competition emerged from HPFH patients that showed mutations in the γ-globin genes, which caused increased γ-gene expression in adult life. Importantly, this leads to a decrease of β-globin gene expression from the same allele while expression of the non-mutated allele is unaffected [210]. In a 'classical' enhancer assay, a small 244 bp fragment of polyoma virus DNA was found to enhance transcription of a linked rabbit β1-gene, while a complete copy of polyoma virus DNA interfered with the enhancement [211]. Expression of a β-globin transgene was found at all developmental stages when it was directly linked to the LCR [16, 212]. This is further supported by the DNase I hypersensitivity of the promoter regions of all globin genes at early stages of development although transcription of the adult genes is absent [8]. Correct developmental timing of β-globin gene expression was achieved by introducing a γ- or α-globin gene in-between the LCR and the β-gene. This correct timing was not achieved when the β-gene was the globin gene closest to the LCR [213]. This suggests competition of the genes for LCR function based on gene order or relative distance. This competition model was confirmed in the context of the complete locus. Introducing a second 'marked' β-gene (βm), located proximal to the LCR with respect to the β-gene, demonstrated a competitive advantage of the proximal βm-gene in terms of expression levels. When the relative distance was reduced between the genes and the LCR the competitive advantage of proximal βm-gene was decreased [214]. In addition, inverting the gene order of the locus, thereby altering relative distances, activated the β-gene at early stages and abolishes ε -gene expression [215]. At the single cell level, *in situ* hybridization of pre-mRNA transcripts supported the competition mechanism by demonstrating that the β-globin genes are alternately transcribed, i.e. only one gene is transcribed at any given moment [111, 216]. A similar alternating transcription was observed for the mouse β-globin genes [217]. Furthermore, the actual switching between γ and β-gene transcription is a continuous dynamic process (a flip-flop mechanism) in which both genes can be alternately transcribed until the γ-genes are permanently silenced [111]. Recently, a structural basis for this competitive interaction was revealed by showing *in vivo* that actively transcribed mouse adult globin genes physically contact the LCR while the intervening DNA bearing inactive embryonic genes is looped out (see Chapter 3) [190,

218]. Moreover, the LCR most likely operates as a single unit, or holocomplex, in these LCR-gene interactions [190].

Of course, *trans*-acting factors play a crucial role in hemoglobin gene switching. As such, the *trans*-acting environment can favor activation of specific genes regardless of gene order [202]. For instance, complete transgenic loci express the γ-genes much higher than the ε-globin gene in mice [199], suggesting that the *trans*-acting environment in embryonic erythroid cells favors LCR-γ-gene interactions. A protein complex has been purified, called direct repeat erythroid-definitive (DRED), that binds a ε-promoter element *in vitro*, which is required for gene silencing [208]. This is a 540 kDa complex and the core consists of two nuclear orphan receptors (TR2 and TR4) that can bind to direct repeat sites in the ε and γ-promoters. DRED was suggested to act as a stage-specific negative regulator by repressing embryonic and fetal transcription in definitive erythroid cells [209]. In addition to the possible role of DRED, EKLF plays a key role in β and γ-gene competition during hemoglobin gene switching. The fetal livers of mice with a reduced EKLF concentration display a decrease in the number of transcriptionally active β-genes with a reciprocal increase in the number of transcriptionally active γ-genes, although the timing of γ globin gene silencing remains the same. When EKLF is absent, there is a further increase in the number of transcriptionally active γ-genes, while β-gene transcription is abolished [181].

1.4.3 Communication between *cis*-regulatory DNA elements in the β-globin loci

As already mentioned earlier gene loci comprise an extensive *cis*-regulatory system and all elements appear to contribute to proper transcriptional regulation. Although many enhancer-promoter combinations are possible, correct enhancer-promoter specificity is observed [4, 10, 20]. This is also the case in the β-globin loci, where the LCR does not influence the expression of the adjacent OR genes [39, 161]. Thus, *cis*-regulatory DNA elements must have mechanisms to communicate with each other in order to drive expression of the correct set of genes, in the right tissue, at the proper time in development.

Hypothetical mechanisms

In the past, several hypothetical mechanisms have been proposed that may explain the communication between the β-globin genes and the LCR (figure 1.4.5). The first model is based on direct interactions between enhancer and promoter sequences. Proteins bind at widely separated enhancers and promoters in order to regulate gene expression and the intervening DNA loops or bends to allow protein-protein interactions [219]. This mechanism is generally known as the looping model. To ensure activation of the β-globin loci by looping a mass-action mechanism has been proposed (reviewed in [220]). In such a system, a transcription factor occasionally binds its cognate binding-site in inactive chromatin when such a binding site is (temporarily) available. Subsequently, two events may occur either the factor dissociates or, within small window-of-opportunity, a second binding site becomes available as the result of chromatin alterations and a second factor binds at such a neighboring site. This process would repeat itself and be the start of inducing a chain-reaction resulting in activation. The significance of the LCR in this mechanism is that it would contain a sufficiently large number of binding sites for factors to ensure activation would usually take place [220, 221]. This simple model incorporates the properties of nucleosome sliding/remodelling [222] and shows similarities with the stop-and-go properties of *trans*-acting factors exchanging between chromatin and the nucleoplasm [150]. Looping to other sites in a locus would lead to chromatin modifications in a larger domain.

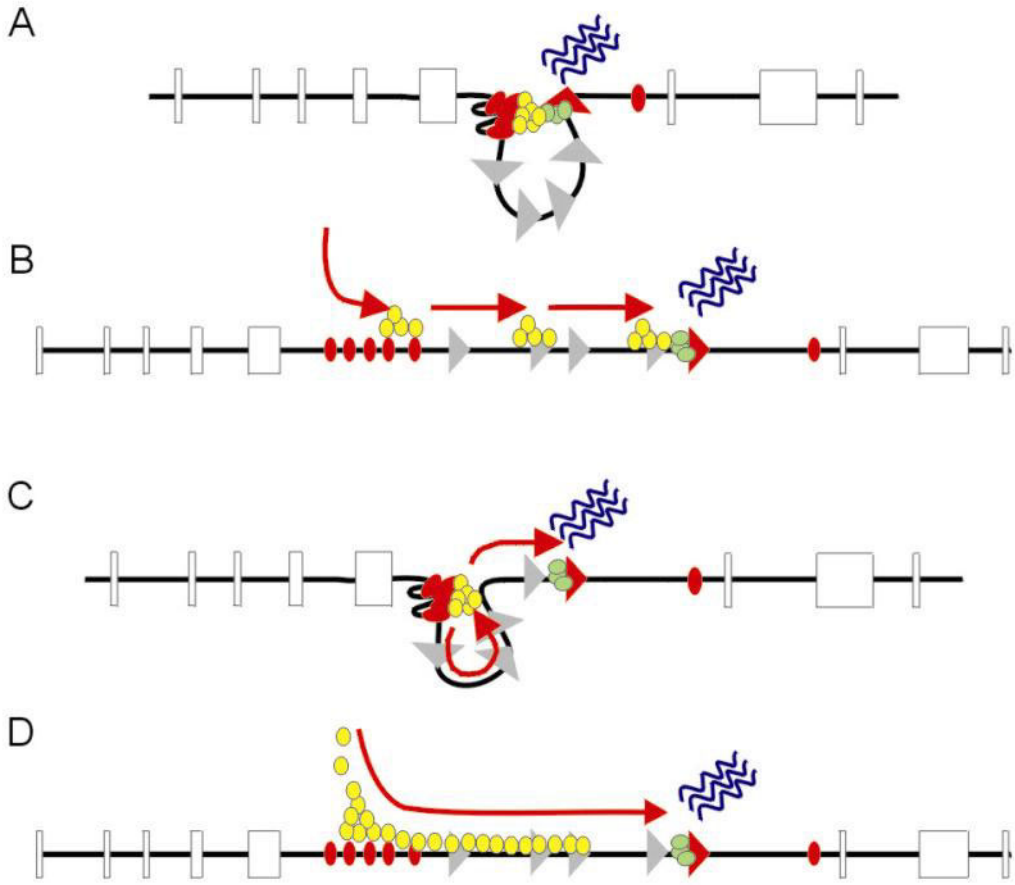


Figure 1.4.5. Hypothetical mechanisms of LCR-globin gene promoter communication.

Globin genes are depicted as triangles with the transcribed gene in red and the nontranscribed genes in grey. DNase I HSs are indicated as red ovals. LCR bound *trans*-acting factors are shown as yellow ovals and promoter bound factors (e.g. general transcription machinery) as green ovals. White boxes are olfactory receptor (OR) genes. The looping model states that DNA-bound protein-protein interactions result in direct contacts between a LCR 'holocomplex' and the promoter of a gene (A). The contacts are established by random-collision and are essential for transcription. As a consequence of the LCR-gene interactions the intervening chromatin template is looped out. The tracking model assumes that factors nucleate at the LCR, which is followed by tracking of those factors to the promoter of the gene (B). The red arrows indicate the movement of the complexes, which depends on the intervening chromatin template. Once a tracking complex reaches the promoter of a potentially active gene then transcription takes place. The facilitated tracking model unites looping and tracking (C). Complexes nucleate and stay bound at the LCR. This LCR-protein complex tracks through the intervening chromatin template towards a potentially active gene (red arrows indicate movement). Upon direct contact between the LCR-protein complex and the promoter transcription occurs. The linking model presumes that LCR-promoter communication is facilitated by transmission of a nucleoprotein structure along the intervening chromatin template (D). This structure originates from the LCR and gradually polymerizes in the direction of the promoter of a gene (red arrow). Transcription happens when the nucleoprotein structure reaches a promoter of a potentially active gene. Note that the looping model, unlike the other models, does not depend on the presence of the intervening chromatin template for transcription, rather transcription depends on random-collision between LCR-bound complexes and promoter-bound complexes.

A second model assumes that enhancer-promoter communication involves proteins that move or track from the enhancer to the promoter using the intervening DNA. This so-called tracking model was supported by the observation that intergenic transcripts were found originating from the 5' HS2 enhancer of the LCR [223]. Similar intergenic transcripts were observed in bacteriophage T4 late genes [224]. This was explained as the basal transcription machinery tracking from enhancer to promoter and consequently long strands of RNA through the intervening DNA are produced.

A third model combines the features of looping and tracking and is therefore referred to as facilitated tracking. As such, an enhancer bound complex 'tracks' through small steps (and perhaps tracking) along the intervening chromatin until it encounters the cognate promoter, at which a stable loop is formed [4]. This model was based on the assumption that the likelihood of forming small loops seems to be higher than the chance of forming large loops, which would occur in the looping model (see box: chromatin contacts). Recent data provided supportive evidence for this model [48]. Chromatin immunoprecipitation (ChIP) analyses at the HNF-4α locus during CaCo-2 cell differentiation demonstrated an independent initial assembly of factors at the enhancer and promoter elements. Subsequently, enhancer-bound factors were present at the intervening DNA, but not upstream of the enhancer, whereas promoter-bound factors stayed fixed at the promoter. Transcription initiation was observed only after the enhancer-bound factors were also detected at the promoter. Detection of the enhancer-bound complex coincided with a unidirectional spreading of histone acetylation [48].

Finally, a fourth model supposes that enhancer-promoter communication needs to be facilitated by transmission of specific nucleoprotein structures from enhancer to promoter; this is referred to as the linking model [225, 226]. Linking was proposed to overcome the prohibitive effects of increasing distance. In addition, spreading of nucleoprotein alters (opens) chromatin structure of intervening DNA and boundary elements determine amount of spreading, resulting in distinct chromosome domains. This model was largely based on the Chip protein in *Drosophila*, which has been suggested to act as a general facilitator of enhancer-promoter communication [225]. Initially it was proposed that enhancer and promoter communicate without directly contacting each other, however, recent discoveries by others necessitated a revision that combined linking and looping in one model [227].

Box: chromatin contacts

The efficiency of bringing chromatin bound-protein factors together by looping depends on the flexibility and conformation of the intervening chromatin template. The stiffness of a chromatin fiber is thought to be determined by interactions between nucleosomes. Thus, if the attractive forces between nucleosomes are relatively weak, the fiber will be soft and bend easily. Based on fluorescence *in situ* hybridization (FISH) data and computational analysis the separation distance of any two given sites that is optimal for interaction can be calculated. For the observed average chromatin fiber optimal site separation is 40 kb with a local concentration of the sites of $8x10^{-9}$ M. However, if the flexibility is increased ~5-fold, presumably due to an open chromatin configuration of the fiber, than the optimal site separation decreases to 10 kb, while the local concentration increases ~100-fold. Local concentrations drop for site separation distances larger or smaller than the optimal site separation, thus the likelihood of interaction will go down. In metaphase chromosomes interactions between distant sites become rather unlikely except for sites that are brought into proximity by the condensation process itself. This effect might also occur within interphase chromatin, for instance as a result of higher-order chromatin folding.
(For details see: [228])

Direct interactions between *cis*-regulatory DNA elements and loop formation

Recently, two independent novel assays demonstrated that the *cis*-regulatory DNA elements of the mouse and human β-globin loci are spatially close and may communicate by interacting with each other in the nucleus of erythroid cells, but not in non-erythroid cells [190, 191, 218]. One of those novel techniques involved RNA TRAP in which HRP-labeled probes are targeted to nascent RNA transcripts and HRP-catalyzed biotin deposition on nearby chromatin is quantified. Clearly, a peak of biotin deposition at 5' HS2 in addition to a peak at the transcribed gene was observed [218]. This finding indicates that the classical enhancer element 5' HS2 of the LCR is in close physical proximity to an actively transcribed β-globin gene *in vivo* [218]. RNA TRAP is dependent on nascent transcripts coming from actively transcribed genes, which can be both an advantage and a disadvantage. The advantage is that nonexpressing cells do not contribute to the measured values, meaning that interactions within the transcriptionally active locus can be picked up even in mixed cell populations. However, to determine whether interactions are specific for the transcriptionally active status of a locus one needs to know its silent conformation, which is an important control that cannot be checked by RNA TRAP. Moreover, it may be difficult to obtain satisfying results by RNA TRAP for genes with moderate transcription rates. The globin genes, which were used to develop RNA TRAP, are transcribed very efficiently, and these genes are exceptionally good targets for visualizing nascent transcripts. The other approach was initially developed in yeast to detect the frequency of interaction between any two genomic sites, which can be used to analyze the overall spatial organization of chromosomes. This methodology is referred to as Chromosome Conformation Capture (3C) [229]. In brief, this method uses formaldehyde to induce protein-protein and protein-DNA cross-links, cleavage of the cross-linked chromatin by a restriction enzyme, intramolecular ligation, and quantitation of ligation products (see chapter 3 and 4) [190, 191, 229]. In potential, 3C can measure cross-linking frequencies at many positions in the locus (e.g. active gene, LCR, intergenic sequence) with the position of restriction sites as the only limitation. Additionally, it can measure cross-linking frequencies in any tissue either expressing a particular gene or repressing it. However, interactions between distal DNA elements are thought to be dynamic, while the measurements represent steady-state average levels. Thus, short-lived but important interactions may give much lower cross-linking frequencies than more long-lived interactions. Furthermore, interpretation of 3C data in a strictly quantitative manner may also be influenced by several other observations. In the first place, we find that additional parameters, e.g. the fragment size, notably affect the cross-linking efficiency. Secondly, changes in fixation conditions differentially affect cross-linking frequencies. Thirdly, it is impossible to say if chromatin is folded as a 10nm fibre, a 30nm fibre or that it adopts yet another conformation. Moreover, chromatin folding is certainly not uniform along the entire β-globin locus, given the presence of DNase I HSs [230]. Therefore, we currently prefer to describe data obtained by 3C technology in a qualitative manner, rather than interpreting the relationship between cross-linking frequency and genomic site separation in terms of real distances. Measured interactions become particularly meaningful if they can be correlated to a phenotype, e.g. if they occur only in a transcriptionally active locus.

Chapter 2

Visualization of the transcriptional dynamics of the human β-globin locus in living cells.

Visualization of the transcriptional dynamics of the human β-globin locus in living cells.

Bas Tolhuis, Stefania Bottardi, Eric Milot, Wouter de Laat, and Frank Grosveld

Department of Cell Biology and Genetics, ErasmusMC, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Fax: +31-10-4089468

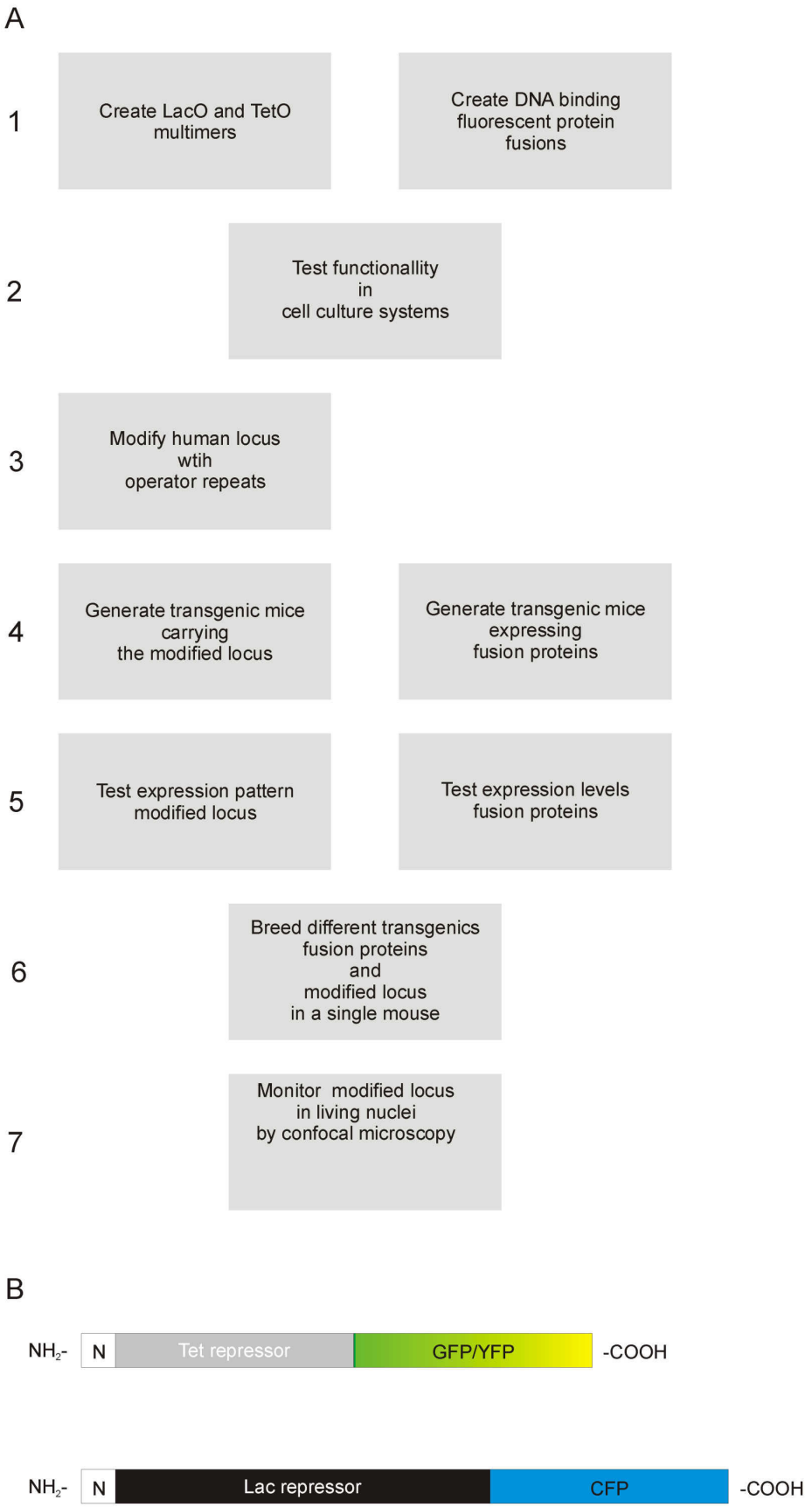
Summary

The human β-globin LCR is a set of upstream regulatory elements that communicates with globin gene promoters to drive transcription. The looping model postulates that communication between the LCR and gene promoters depends on direct contact between LCR-bound and promoter-bound protein complexes. As a result, intervening chromatin would be looped out. These interactions are likely to be dynamic rather than static, as it has been shown that globin genes are alternately transcribed [111]. Here, we describe an approach to 'mark' the human β-globin locus with a dual tagging system, using Lac^O/LacR-cyan fluorescent protein (CFP) in conjunction with a Tet^O/TetR-yellow fluorescent protein (YFP). This tool should allow monitoring of dynamic interactions between the communicating elements. This work is still in progress.

Introduction

The development of a Lac^O/LacR tagging tool uniquely has allowed examining the dynamics of the chromatin template in living cells [135]. This technique was based on the insertion of 256 direct repeats of the bacterial Lac^O sequence into the genome of cells. Subsequently, cells having undergone gene amplification were selected using dihydrofolate reductase (DHFR) as a selectable marker. The large tandem arrays (tens of Mbp in size) of the Lac^O sequence in the resulting cell lines were visualized by expressing a fusion gene between green fluorescent protein (GFP) and LacR genes. The latter is the cognate DNA binding protein of Lac^O and targets GFP to the tandem arrays. This technique is particularly useful in visualizing chromatin regions in living interphase nuclei, which are otherwise undetectable due to the decondensed state of interphase chromatin.
With the Lac^O/LacR tagging tool, several observations were made that provide insight in the dynamic nature of chromatin regions. First, regions in yeast, Drosophila, and mammalian cells show a constrained diffusion rate that consists of small (~0.5 μm radius) and rapid (1-10 seconds) random walk movements [90, 136-138]. ATP-depletion of yeast cells largely suppressed the random walk movements, whereas destabilizing microtubules did not affect movement [136]. Therefore, the movement was interpreted to be driven by large ATP-dependent enzyme activities involved in transcription or chromatin remodeling [90]. In addition, nuclear processes (e.g. DNA replication), chromatin state, and nuclear compartments can influence the mobility of a particular chromatin region [90, 136, 137]. Secondly, a slow (30-60 minutes) long-range motion was observed in *Drosophila* that is confined to a much larger, chromosome-sized domain (~3 μm radius) [138]. Thirdly, the morphological structure of large chromatin regions (~80 Mbp Lac^O arrays) showed either a fibrilar or a solid density LacR-GFP staining. Electron microscopy showed that the fibrilar structures were 80-130 nm in diameter, suggesting that chromatin folding may well exist above the 30-nm chromatin fiber-packaging ratio in living interphase nuclei [135]. Fourthly, solid density stained arrays decondensed dramatically when the potent VP16 transcriptional activation domain was targeted to the Lac^O sequences [139, 140]. Furthermore, it was demonstrated that the decondensation was not merely a consequence of ongoing transcription [139]. Instead, it was accompanied by histone hyperacetylation and recruitment of histone acetyltransferases and chromatin remodeling enzymes [139, 141]. Thus, the transcriptional activator VP16 dramatically modifies large-scale chromatin structure presumably by recruiting enzymes that alter nucleosome structure.

Theoretically, this tagging tool should also be suitable to study the dynamic interactions between two or more genomic sites. In particular, it would be interesting to study the dynamics of transcriptional regulatory elements that are known to communicate with each other, but that are located to sites far apart on the linear DNA template. The human β-globin locus contains such separated regulatory elements that need to communicate to drive transcription of the globin genes. As


A
1
Create LacO and TetO multimers
Create DNA binding fluorescent protein fusions
2
Test functionallity in cell culture systems
3
Modify human locus wtih operator repeats
4
Generate transgenic mice carrying the modified locus
Generate transgenic mice expressing fusion proteins
5
Test expression pattern modified locus
Test expression levels fusion proteins
6
Breed different transgenics fusion proteins and modified locus in a single mouse
7
Monitor modified locus in living nuclei by confocal microscopy
B
NH2-
N
Tet repressor
GFP/YFP
-COOH
NH2-
N
Lac repressor
CFP
-COOH

described in chapter I (paragraph 1.4.1) the locus spans a region of approximately 70 kb containing five developmentally regulated globin genes arranged in the order 5'-ε, ^{G}γ, ^{A}γ, δ, β-3'. Upstream of the ε-globin gene a set of developmentally stable DNase I HSs are present that are referred to as the locus control region (LCR) [16, 163]. In two human γδβ-thalassemias large deletions of the locus were observed. One deletion removed the LCR and all genes ,but the β-globin gene and its flanking regions were still intact (Dutch). The other deletion removed most HSs of the LCR, namely 5' HS2 to 5' HS5 (Spanish), while all the genes were still intact. In both cases, this resulted in both cases in abrogated expression of all gene(s) and an altered chromatin state in the Spanish deletion [23, 165, 192, 198]. In transgenic mice, the LCR is sufficient to drive position independent and copy number dependent expression of a linked gene [16]. Thus, the LCR appears to be a set of upstream regulatory elements that communicates with globin gene promoters to drive transcription.

Transcriptional regulation in the locus is based on autonomous silencing of particular genes and competition between genes that depends on their relative distance to the LCR [19, 213]. LCR proximal genes show a competitive advantage over distal genes [213]. In addition, reducing the relative distance between two genes and the LCR reduces the competitive advantage of the proximal gene [214] and reversing gene order in the locus results in embryonic expression of the β-globin gene that is now proximal to the LCR. In addition, expression of the ε-globin gene was undetectable, since it was placed distally by the inversion [215]. A model that explains the mechanism by which proximal genes can suppress those located more distally involves direct contacts between LCR and gene promoters and consequently looping of intervening DNA sequences. Stage-specific negative regulators acting on sequences in the ε- and γ-gene promoters autonomously silence these genes [158, 204-206], allowing transcription of downstream located genes.

Detection of primary globin transcripts *in situ* suggested that globin genes are transcribed alternately rather than at the same time [111, 216, 217] and transcription can switch back and forth between genes in a flip-flop mechanism [111]. These data imply that communication between LCR and gene promoters is a dynamic rather than static process. It would be of great interest to confirm and expand these data by visualizing the communication between the LCR and the β-globin gene promoter in living cells and study the dynamics of this communication. Here, the proof of principle is described of an approach to 'mark' the human β-globin locus with a dual tagging system that consists of the Lac^{O}/LacR-cyan fluorescent protein (CFP) in conjunction with a Tet^{O}/TetR-yellow fluorescent protein (YFP).

Figure 2.1. Schematic outline of visualizing the locus and the fluorescent fusion proteins.

The strategic setup of the dual tagging tool is shown (A). First, we make several DNA constructs either multimers of operator sequences or DNA-binding fluorescent fusion protein expression modules (1). Next, we need to establish the minimal number of detectable binding sites and fusion protein functionality in mouse erythroleukemia (MEL) cells (2). The optimal number of binding sites will be inserted into a PAC clone containing the complete human β-globin locus by homologous recombination in *E. coli* [232] (3). On each side of the locus a operator repeat will be inserted, Tet^{O} 5' and Lac^{O} 3'. Transgenic mice are generated that carry the modified human locus or fusion protein expression modules (4). The expression of the globin genes in the modified locus are analyzed and the expression levels of the fusion proteins are determined (5). Extensive breeding will generate mice that carry a functional modified human β-gobin locus and two different fusion protein modules that express the transgenes at appropriate levels (6). Finally, time-lapse confocal microscopy will be used to monitor the dynamic interactions between the two ends of the locus (7). The fusion proteins involved are presented schematically (B). Each contains an N-terminal nuclear localization signal (N) followed by the DNA binding domain and a fluorescent protein. The DNA-binding domains used are Tet repressor (TetR, binds Tet$^{O)}$ and Lac repressor (LacR, binds Lac^{O}), indicated in grey and black bars respectively. The TetR was fused to either GFP or YFP (green/yellow bar) and LacR was fused to CFP (blue bar).

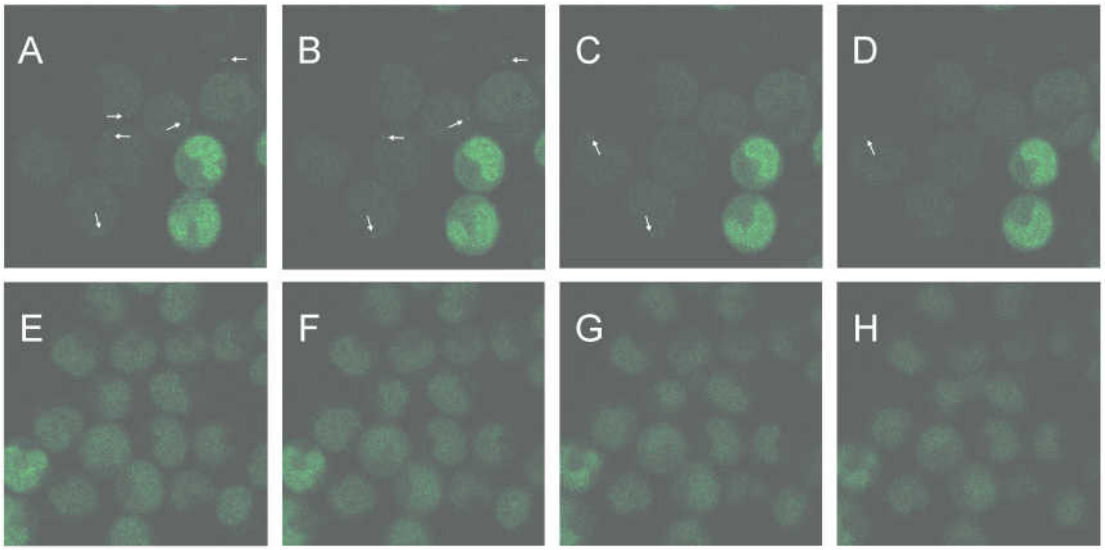


Figure 2.2. A 28-mer of the Tet^O binding sequence can be detected in living cells by confocal microscopy.

Four consecutive optical sections are showing 28xMEL Tet^O cells (A-D). DNA loci are visualised through bound TetR-GFP proteins that are indicated by white arrows. Note that expression levels of TetR-GFP differ per cell, and that visualization of the DNA locus requires low levels of the fusion protein. A control MEL cell line expressing TetR-GFP shows no obvious signals (E-H).

Results

Only large tandem arrays of the Lac^O binding sequence have been analysed. The sizes of these arrays range from 90 Mbp to almost 10 kbp (corresponding to ~3x10^6 and 256 Lac^O sequences, respectively) [135, 139], which is unsuitable for analyzing the transcriptional dynamics of the human β-globin locus. First of all, a 90 Mbp tandem array is almost a 1000-fold larger than the human locus, hence the dynamics of the locus might be insignificantly small compared to the dynamics of the complete array. Secondly, insertion of DNA sequences in-between LCR and globin genes will alter the relative distance and may affect transcriptional regulation. Thirdly, one report observed co-localization of the operator array to PML bodies upon the binding of the LacR-GFP. This was interpreted as local accumulations of foreign factors or DNA that are detected and marked by the 'sensor' PML body [140]. It would therefore be necessary to keep the insertion of repeat sequences in the human locus as small as possible. However, the signal-to-noise-ratio should be high enough to be able to detect the operator sequences in the entire nucleus. Therefore, the minimal number of detectable binding sites by fluorescent microscopy needed to be established. Figure 2.1A gives a schematic outline of the experimental approach.

First, multimers of the Tet^O and Lac^O sequences were constucted followed by the fusion proteins that consisted of a DNA binding protein (either TetR or LacR) and a fluorescent protein (either GFP or one of its derivatives) (see Methods and figure 2.1B). Next, we prepared stably transfected mouse erythroleukemia (MEL) cells with a construct bearing a 7-mer Tet^O repeat (~0.3 kb) and a neomycin selectable marker. Specific MEL clones were isolated and analysed by Southern blot to determine the copy number of the integrated 7xTet^O-construct (data not shown). One clone contained four integrated copies of the 7xTet^O-construct, designated MEL 28xTet^O that was used in a second transfection round to introduce a TetR-GFP fusion protein. Confocal microscopy was used to analyze

the MEL 28xTet^{O}-TetRGFP cell population. Figure 2.2A-D shows four sequential confocal images of several cells of this population and indicates that in a subset of the cells small spot can be observed. As a control TetR-GFP in MEL cells without Tet^{O} sequences were introduced (figure 2.2E-H) and no obvious spots could be detected. The number of spot containing cells were quantified by analyzing 100 MEL 28xTet^{O}-TetRGFP cells, out of these 35 cells showed a single spot, whereas in the control cells (n=153) only 3 cells contained a single spot. Interestingly, we found that that most spot-containing 28xTet^{O} cells exhibited low background-fluorescence intensity. For instance, the intensities of the spot-containing cells in figure 2.2A-D is much lower than the two cells at the right bottom of the images that do not show any spots. Expression of the TetR-GFP fusion was driven by an erythroid cell-specific microlocus expression cassette [231] with a mutated CAAT element in its β-globin promoter (see Methods). This mutated promoter expresses a linked gene at 57% compared to the wild-type β-globin promoter. (Ernie de Boer, personal comm.). The cells were not cloned after stable transfection with this cassette. Instead, a pool of cells with different integration sites and copy number of the cassette were used, enabling an analysis of various TetR-GFP expression levels. Thus, the varying fluorescence intensities most likely reflect distinct expression levels of the TetR-GFP fusion protein. This led to the conclusion for optimal signal-to-noise-ratio to visualize a low number of binding sequences, such as in the MEL 28xTet^{O}, the nuclear concentration of its fluorescently tagged cognate binding protein has to be low to reduce the background-fluorescence. Other fusion constructs (e.g. LacR-CFP and TetR-YFP) were also tested for functionality in cell culture systems, they all showed

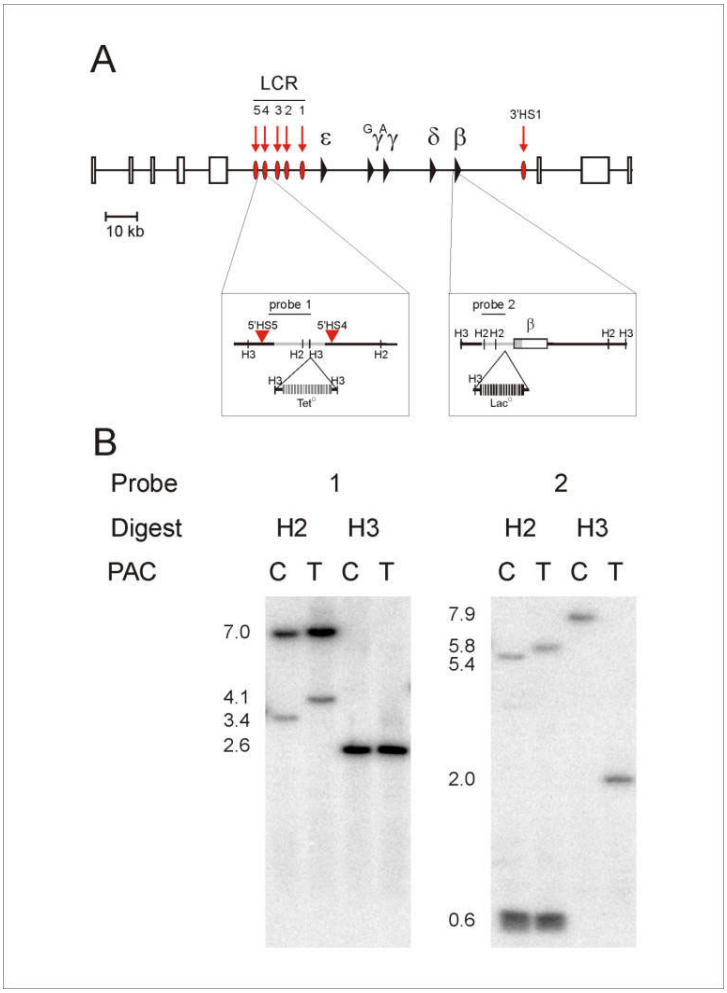


Figure 2.3. Tet^{O} and Lac^{O} sequences in PAC TL-O.

A schematic presentation of the PAC148γ-lox is shown (A). The locus contains the globin genes (black triangles), DNAse I HSs (red arrows) and olfactory receptor genes (white boxes). The homologous recombination strategies are pointed out (boxes). The region used for homologous recombination is 'marked' in grey and the operator insertion positions are specified. The β-globin gene is visible as a white box and 5'HS5/5'HS4 as red arrowheads. Restriction sites used in southern blot analysis are shown (H2: *Hind*II; H3: *Hind*III). Southern blot analysis with probe 1 (B left panel, see Methods) gave as expected a shift in a *Hind*II digest compared to a control PAC. The *Hind*III digest did not show a shift, because the probe cannot detect the 14xTet^{O} *Hind*III fragment. The same analysis with probe 2 (B right panel, see Methods) gave an unexpectedly low size shift with a *Hind*II digest (see text). However, the Lac^{O} repeat was inserted, since an additional *Hind*III site reduced the detected fragment size. C and T point out control PAC and PAC TL-O, respectively. Sizes are indicated (kb).

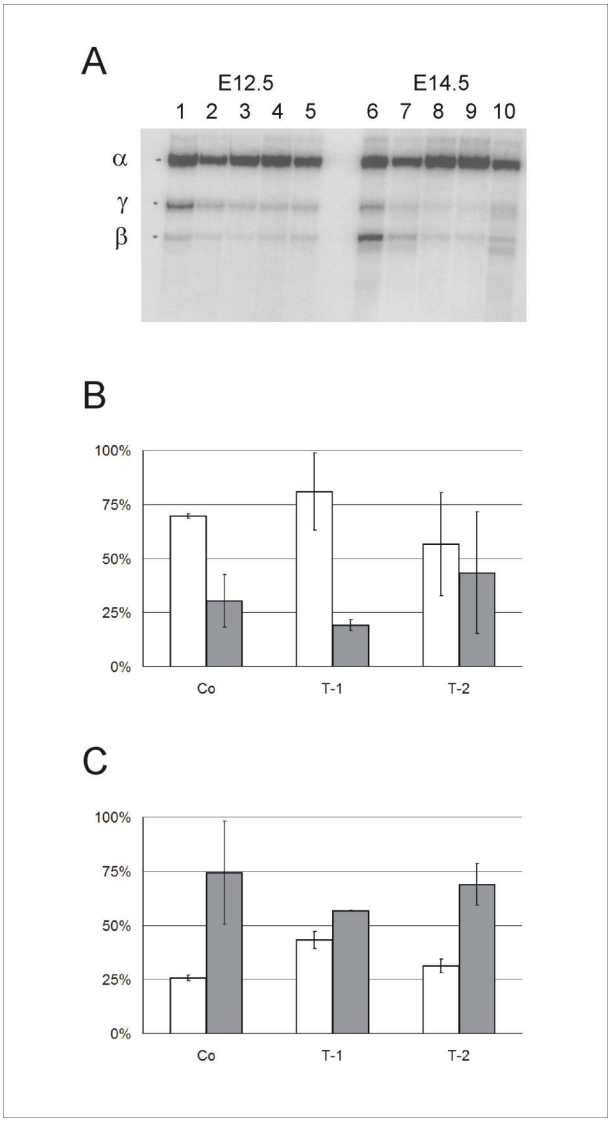


Figure 2.4. Transcription levels in PAC TL-O transgenic mice.

An example of an S1 Nuclease protection assay (A) of ~1 μg 12.5dpc fetal liver RNA (lanes 1-5) and ~1 μg 14.5dpc fetal liver RNA (lanes 6-10). Lanes: 1 and 6, 3x control PAC; 2 and 7, 1x control PAC; 3 and 8, PAC TL-O line 1 (sample A); 4 and 9, PAC TL-O line 1 (sample B); 5 and 10, PAC TL-O line 2. The samples were assayed with radioactive labeled probes for mouse α, human γ and β. Corresponding protected fragments are indicated. Quantification of all S1 Nuclease protection experiments shows that ratios between human γ and β-genes are comparable between the control (Co) and the PAC TL-O (T-1 and T-2) lines at 12.5dpc (B). The PAC TL-O 1 line (T-1) appears to have slightly inreased γ-globin levels at 14.5dpc (C), while Pac TL-O 2 line appears normal. For quantification mouse α was used as a loading control and the graphs are ratios of (γ or β)/(γ and β). The cumulative control PAC values set to 100%. The data represent averages of ~3 independent S1 Nuclease protection assays and standard errors are indicated. The white bars represent γ-globin levels and the grey bars β-globin.

fluorescence and nuclear extracts from a LacR-CFP cell lines were capable of binding a Lac^O oligomer *in vitro* (see figure 2.5).

Next, we introduced Tet^O and Lac^O repeat sequences at the 5'-end (LCR) and the 3'-end (β-globin promoter) of the locus, respectively. For this, recently developed homologous recombination strategies in *E. coli* have been applied, using a PAC containing 185kb of human DNA including the complete globin locus [232]. A ~3.2kb *Eco*RV/*Stu*I fragment of the human β-globin LCR with a 14-mer Tet^O repeat (~0.6 kb) inserted into a *Hind*III site was used for homologous recombination, which placed the 14xTet^O between 5' HS5 and 5' HS4 (figure 2.3A). The 5' HS5 does not seem to contribute to activation of transcription of the globin genes [35] and at this position the repeat will be part of the putative LCR 'holocomplex' [28, 180, 182]. Southern blot analysis of the PAC clone with a probe located just upstream of the *Hind*III integration site, gave a diagnostic shift in a *Hind*II digest compared to a control PAC (figure 2.3B). Subsequently, we used a ~1.9 kb *Bam*HI fragment with a 32-mer Lac^O (~1.6 kb) sequence inserted into a *Bsm*BI site, placing the insert ~400 bp upstream of the RNA start site (figure 2.3A). However, a *Hind*II digest showed only a 400 bp shift compared to the control PAC (figure 2.3B). Close examination of the homologous recombination vector revealed that multiple smaller fragments were present (data not shown), suggesting that the plasmid had been rearranged.

Several attempts to introduce repeats longer than the 14-mer Tet^O failed in the recombination vector. Thus, the presence of these repeats may result in rearrangements perhaps in combination with the RecA protein that is expressed from the recombination vector [232]. Once integrated into the β-globin locus (in absence of the recombination vector) the repeats were stable and no additional rearrangements were detected in the modified PAC Tet/Lac Operator (TL-O) (data not shown).

The effect of the small repeats on the transcriptional regulation of the β-globin locus was tested subsequently. The PAC TL-O was used to generate transgenic mice and two different lines were obtained that showed integration of intact construct as determined by Southern blot (data not shown). They contained 3 and 4 complete copies. We used S1 nuclease protection to examine the globin transcript levels of the PAC TL-O transgenics compared to a control double copy PAC transgenic (figure 2.4) [233]. Quantification of the data revealed that the expression level of the double-copy control PAC seems to be similar to the multi-copy PAC TL-O 2 line mice, while this is not the case for the PAC TL-O 1 line (table 2.1). At day E12.5 of gestation the γ-genes are expressed at a higher level than the β-globin gene in both the control and PAC TL-O transgenics, while at day E14.5 the β-gene expresses higher in all cases (figure 2.4A). This indicates that the ratio between the β and γ-genes is similar between the control PAC and the PAC TL-O transgenics (figure 2.4B and C), suggesting that developmental stage-specific transcription of the globin genes is properly regulated. However, ε and γ-globin transcription levels in primitive erythroid cells still have to be analyzed to confirm proper developmental regulation. In addition, the PAC TL-O 1 line appears to have slightly reduced β-globin levels at E14.5 and increased γ-globin levels (figure 2.4C).

In addition to PAC TL-O transgenic mice, we also generated transgenic mice carrying expression cassettes containing TetR-YFP, LacR-CFP or both. We obtained 17 independent transgenic lines (9 TetR-YFP, 5 LacR-CFP, and 2 TetR-YFP/LacR-CFP double transgenics). Initially, we used fluorescence microscopy to check for expression levels of the fusion proteins, since the intensity of fluorescence is critical for visualization of operator repeats. However, fluorescence emanating from the fusion proteins was not observed in 12.5 dpc fetal liver cells or nucleated adult blood cells from anemic animals (data not shown). On the other hand, we could observe GFP fluorescence driven by an actin promoter in fetal liver cells, suggesting that the functionality of GFP is not impaired in erythroid cells (data not shown). Nuclear extracts from fetal livers (12.5 dpc) of a LacR-CFP transgenic line was prepared and analyzed by gel mobility shift assay (figure 2.5) and Western blot analysis (data not shown). In contrast to control extracts obtained from cell lines, no oligomer shift or protein could be detected. This suggested that the protein was not detectable in nuclei of the transgenic animals. The data from the fusion constructs are strikingly different between cultured cells (figure 2.2) and cells derived from transgenic animals (figure 2.5). In cultured cells, the fusion proteins

Table 2.1. Total expression levels of the PAC TL-O human γ and β-globin genes relative to a control PAC transgenic.

Time point	Control PAC	PAC TL-O 1	PAC TL-O 2
12.5dpc	1.00 ±0.13	0.22±0.02	0.80±0.21
14.5dpc	1.00±0.19	0.08±0.00	0.64±0.03

The expression levels of human γ and β-globin were corrected for copy number and the cumulative control PAC values were used to normalize to 1. The data represent averages of ~3 independent S1 Nuclease protection assays and standard-errors are indicated.

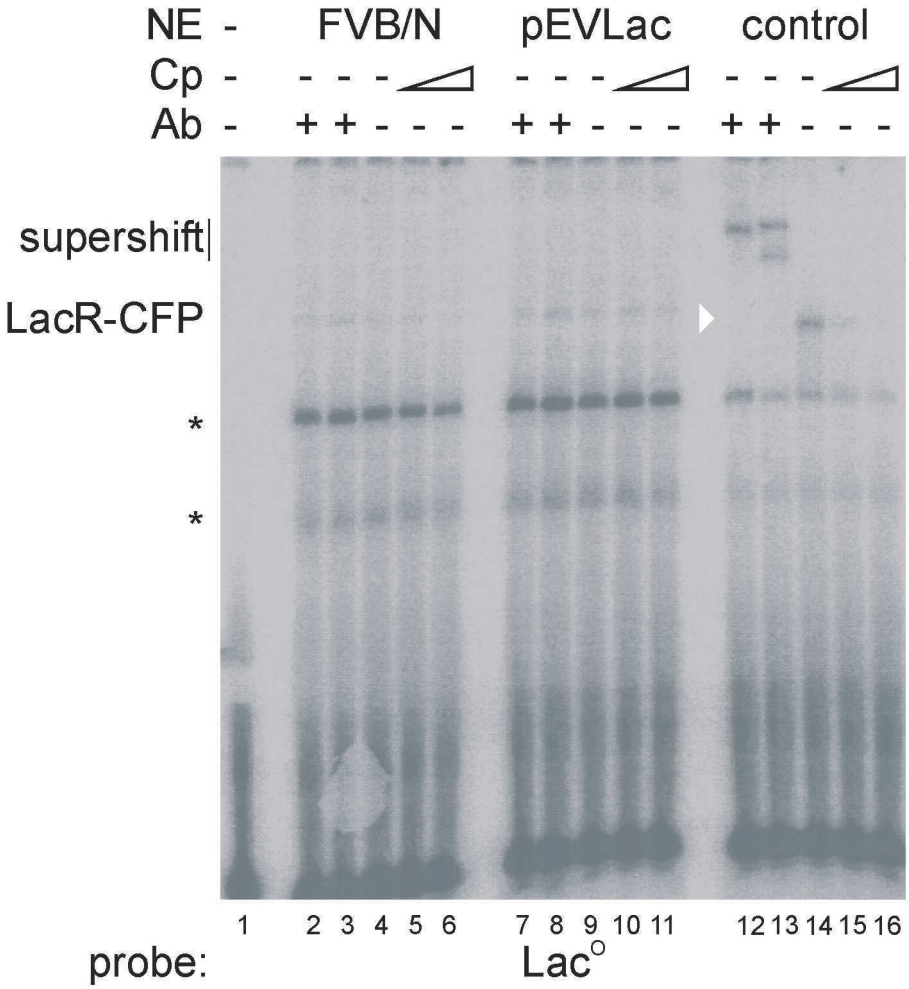


Figure 2.5. Gel mobility shift assay.

Nuclear extracts (NE) were made from wild-type (FVB/N) or LacR-CFP transgenic (pEVLac) 12.5 dpc fetal liver cells and MEL cells transfected with a LacR-CFP construct (control). Lanes 4, 9, and 14 were incubated only with Lac^O probe. A shift in mobility of the probe is observed in the control extract (LacR-CFP binding is indicated by white arrowhead, lane 14) but not with the wild-type and transgenic extracts. The observed shift in the control can be competed by addition of 10-fold (lane 15) and 100-fold (lane 16) molar excess of cold competitor Lac^O (Cp). This competition cannot be observed in wild-type and transgenic extracts (lanes 5, 6, 10, and 11). Supershifts with GFP antibodies (Ab), either a monoclonal α-GFP (lane 12) or a polyclonal α-GFP (lane 13), can only be observed with the control extract. Aspecific binding of the Lac^O probe occurs in all extracts (indicated by asterisks) but can only be efficiently competed by specific binding in the control extracts (lanes 12-16).

are expressed and functional, while this is not the case in transgenic animals. Hence alternative methods to express the fusion proteins in mouse cells carrying the modified PAC TL-O locus are currently under development (see discussion).

Discussion

We have attempted to develop a tool to monitor the dynamic chromatin interactions in living cells. The strategy was to 'mark' the human β-globin locus with a dual tagging system, using Lac^O/LacR-CFP in

conjunction with a Tet^O/TetR-YFP. We demonstrated that a 28-mer of Tet^O sequences was detectable in living MEL cells. Furthermore, we conclude that a low expression level of the TetR-GFP fusion protein was essential to obtain an optimal signal-to-noise-ratio. This is in agreement with previously described experiments with larger tandem arrays of Lac^O detected by LacR-GFP [135]. The two tags were positioned at distinct positions in the locus by homologous recombination in *E. coli*. One transgenic line carrying the modified human locus showed that the inserted operator repeats did not interfere with transcription of the β-globin genes. The other transgenic line showed reduced expression levels that may be explained by either variegated expression or reduced transcription rate per locus may be the cause of the reduced expression levels. mRNA-fluorescent *in situ* hybridization (FISH) could be used to distinguish between the two possibilities [28]. In addition, a slight difference in developmental gene regulation was observed in this line, since the γ- to β-gene expression ratio at E14.5 seemed to be higher than that of the control cells. It is known that perturbations at the β-promoter that reduce β-globin expression coincide with upregulated γ-globin expression [181]. Thus, the insertion of the Lac^O repeat in this construct may interfere to some extent with its expression. This is not important for the primary aim of the experiment. Due to the number of transgenic lines it is not known whether the altered transcription levels in the PAC TL-O 1 trangenics are an exception or wether reduced expression levels of the globin genes in the PAC TL-O locus occur frequently.

The lack of detectable expression in transgenic animals carrying TetR-YFP and LacR-CFP is in great contrast with the expression and functionality of the same constructs in cultured MEL cells. This may be explained by the observation that transgene silencing is often associated with the presence of CpG-rich prokaryotic sequence repeats that make them prone to methylation and silencing (reviewed in [121]). Indeed, transgenic mice containing the *LacI* gene, coding for the LacR protein, showed heavy methylation and in most tissues silencing of the transgene [234, 235]. Recently, the primary DNA sequence of the *LacI* gene was changed to resemble mammalian coding usage more closely but still code for the same protein. This construct did express a functional LacR protein in transgenic mice [236]. A change of codon usage of the *LacI* gene from our constructs is presently in progress. It should be noted however that a number of experiments have successfully used the TetR encoding transgenes to conditionally control gene expression with the Tet-on/off system in the mouse (reviewed in [237]). Thus, the observed lack of expression from our constructs needs further examination to be fully understood, e.g. it is not known whether the reporter constructs are transcribed, and whether they are methylated. One possibility to circumvent methylation would be introducing the reporter construct via viral transduction into the transgenic mouse cells. In a pilot experiment, we introduced both LacR-CFP and TetR-YFP into cultured 12.5 dpc fetal liver cells [238, 239], using Moloney retroviral transduction [240]. This showed that fluorescence could be detected in the fetal liver cells with both constructs (data not shown).

Recent data obtained by our laboratory demonstrated spatial interactions between the LCR, actively transcribed genes and distal DNase I hypersensitive regions occur *in vivo*. Furthermore, the data support the existence of an erythroid cell-specific nuclear compartment dedicated to the transcription of the globin genes by RNAP II, called the active chromatin hub (ACH, see chapter 3 and 4 [190, 191]). The ACH model provides a mechanistic framework that may improve knowledge of transcription in the 3-dimensional space of the nucleus. These interactions were measured using a novel biochemical method, called chromosome conformation capture [190, 229]. However, these measurements represent steady-state average levels. Consequently, the dynamic interactions between LCR and globin genes cannot be detected.

Although the development of the dual-tagging tool is still incomplete, this technology has enormous potential. For instance, it would allow a study of the dynamics of interaction of transcriptional regulation in living cells during development and in differentiation pathways. Additionally, changes in transcription factor environment, such as EKLF over-expression or knockouts,

might give insight into the stability of the dynamic interactions. One important last hurdle to this method may be the microscopy itself. It is presently not known whether the current confocal microscopy techniques can obtain a resolution that is good enough to visualize the two sides of the locus as separate signals in cells that do not express the β-globin gene. This would be important to detect differences in dynamic interactions between expressing and nonexpressing cells. With our current confocal microscope we can obtain a lateral resolution of approximately 200 nm. In addition, with fluorescence *in situ* hybridization the visualization of DNA probes (~40 kb) that were separated by approximately 150 kb of DNA as distinct signals has been successful [221]. The two operator repeats in the locus are separated by ~70 kb, hence detection of individual signals in nonexpressing cells would be more challenging. Alternatively, fluorescence resonance energy transfer (FRET), which has a spatial resolution that exceeds the optical limit of the light microscope could be used to detect spatial proximity of the operator in real time [241, 242].

Methods

Constructs

Generation of the 7xTet^{O}-construct involved insertion of a 310 bp heptamerized tet operator fragment (*Eco*RI-blunt/*Kpn*I) and a pgkPuro (*Hind*III-blunt/*Kpn*I) fragment into pBlueScript KS- (Stratagene). The 7xTet^{O} fragment was derived from plasmid pUHC13-3 [243]. The TetR-GFP fusion construct required insertion of SV40NLS (*Eco*RI/*Xba*I), 618 bp *tet* repressor PCR fragment (*Xba*I/*Sac*II) and GFP (*Sac*II/*Not*I) into a minilocus expression cassette [231] carrying a neomycin selection marker and a β-globin promoter with a mutated CCAAT box (located approximately –80 bp from the transcriptional start). This mutant contains 6 base substitutions (underlined) 5'-ACTACGAC-3', while a wild-type element contains the following sequence 5'-CCAATCT-3'. (Ernie de Boer, personal comm.). The GFP gene and other variants were derived from pEGFP-N1, pEYFP-N1, and pECFP-N1 (Clontech). The *tet* repressor gene was obtained by PCR from Tn10 (Accession #X00694) with appropriate flanking restriction sites. Primers: sense 5'-TCT AGA TTA GAT AAA AGT AAA AGT AAA GTG ATT-3', antisense 5'-AAA CAA CTT AAA TGT GAA AGT GGG TCT-3'. The LacR construct was similar to the TetR constructs, however a 1081 bp *lacI* gene PCR fragment (*Xba*I/*Sac*II) replaced the TetR. The *lac* repressor gene was obtained by PCR on ~ 10 ng of genomic *E. coli* DNA. Primers: sense 5'-GCT CTA GAA AAC CAG TAA CGT TA-3'; antisense 5'-TCC CCG CGG GTG CCC GCT TTC CAG-3'.

Two recombination vectors used in the homologous recombination of the PAC clone were made. The 14xTet^{O} vector contained a ~3.2kb *Eco*RV/*Stu*I fragment of the human β-globin LCR with two tandem 7xTet^{O} repeats (~0.6 kb) inserted into its *Hind*III site. This ~3.8 kb fragment was inserted *Eco*RV/*Stu*I blunt into a *Hind*III site of pDF25 [232]. For the generation of the 32xLac^{O} recombination vector we first amplified a single Lac^{O} to a 32-mer. We used an insert amplifaction strategy analogous to Robinett et al. [135]. Briefly, a sense and antisense oligomer containing a single Lac^{O} were annealed and inserted into a *Sal*I site of pBlueScript SK+ (Stratagene). A subamount of this pLac^{O}1x, including the operator, was digested with *Sac*II/*Cla*I and another amount was digested with *Sac*II/*Acc*I, also including the operator. These two fragments were ligated together resulting in a duplication of the operator sequences. This strategy relied on the compatible sticky ends generated by *Cla*I and *Acc*I, which after ligation could not be recut with either enzyme. Six rounds of this cloning cycle generated a pLac^{O}32x plasmid. The 32xLac^{O} recombination vector (pDF25) consisted of a ~1.9 kb *Bam*HI β-globin promoter fragment with the 32x Lac^{O} sequence inserted *Hinc*II/*Eco*RV into a blunt *Bsm*BI site. The homologous recombination of PAC148γlox has been described before [232]. Lac^{O} oligomers: sense 5'-GCA GCT CGT GTT TAA CAA TAG GCC GAG TGT TAA GGT GTA CAC-3', antisense 5'-TCG AGC ACA AAT TGT TAT CCG CTC ACA ATT CCA CAT GTG ACG-3'.

Probes used for southern blot (figure 2.3B): probe 1 is a ~1.6kb *Sst*I/*Hind*III fragment surrounding 5' HS5 of the human β-globin locus; probe 2 is a ~1.0kb *Bam*HI/*Bsm*BI fragment located upstream of the β-globin promoter.

Cell culture

MEL C88 cells were stably transfected as described before [244]. Selection of clones was done by diluting the cells either 1/100 or 1/1000 and loading 100 μl aliquots in a 96-wells plate under puromycin selection conditions. Southern blot was employed to determine the number of clones in each well and copy number of the construct, using the Tet^O sequence as a probe. Clone 28xTet^O was subjected to a second round of stable transfection with the TetRGFP expression cassette under neomycin selection conditions. All cells were cultured in DMEM 10% FCS and 1% penicilline/streptomycin.

Microscopy

All microscopic images were taken using a Zeiss LSM510 confocal microscope, using a heating chamber at 37°C and a continous CO_2 flow to keep the cells viable. The cells were kept in appropriate medium conditions during image acquisition (see above). Although fetal liver cells and MEL cells are analyzed in suspension, a significant number of cells sets relatively immobile on to the coverslip, allowing image acquisition. Optical sections of 1μm were taken using an Axiovert inverted microscope with Axiovert lens 63 x oil (1.4 n.a.,working distance 90 μm), an Argon gas laser (458, 488, 514 nm), and a filter LP 475nm.

Transgenic mice

The 185 kb PAC TL-O insert was isolated by *Not*I digestion and purified from vector sequences by salt gradient centrifugation, essentially as described by Dillon and Grosveld [245]. Briefly, the digested PAC was layered on top of a 5-25% NaCl gradient and centrifuged at 40,000rpm, room temperature for 50 minutes in a SW41 swing-out rotor. 0.5ml fractions were collected and analysed by agarose gel electrophoresis. Fractions containing only the PAC insert were pooled and dialysed against a large volume of TE (10mM Tris-HCl pH 8.0, 1mM EDTA)/0.1M NaCl for 5 h at 4°C in UH 100-75 dialysis tubing (Schleicher & Schuell). Dialysis was continued overnight at 4°C after replacing the buffer. The PAC insert was concentrated by vacuum dialysis and subsequently dialysed against a large volume of microinjection buffer (10mM Tris-HCl pH 7.4, 0.1mM EDTA) containing 0.1M NaCl in order to protect the high molecular weight PAC insert DNA from shearing during microinjection. The purified PAC fragment was checked for DNA integrity and concentration by pulsed field gel electrophoresis in a 1% agarose gel in 0.25XTAE buffer using a Biometra RotaphorType V apparatus, under the following conditions: 8-2 seconds pulse interval logarithmic ramp, 120-110° rotor angle linear ramp, 200-180 volt logarithmic ramp, rotor speed 6 at 13°C for 21 hours.

*Aat*II/*Asp*718I digestion isolated the 11.2 kb TetR-YFP and11.6 kb LacR-CFP expression cassettes, including the microLCR, mutated β-globin promoter, and a 2.8 kb β-globin gene fragment for efficient mRNA production. The digested fragments were purified from vector sequences by gel electrophoreses and purified with Gelase (Biozym) according to manufacturer's protocol. DNA was further purified using Elutip-d columns (Schleicher & Schuell) followed by ethanol precipitation. The pellet was resolved in microinjection buffer (10mM Tris-HCl pH 7.4, 0.1mM EDTA).

The purified fragments were injected at approximately 0.5ng/μl into the pronucleus of fertilised eggs of FVB/N mice. The injected eggs were transferred into the oviducts of pseudo-pregnant BCBA foster females as previously described [7]. Transgenic founders were identified via Southern blot analysis using as probes the 970 bp BamHI-EcoRI βIVS2 fragment and a 3.3 kb EcoRI fragment containing HS5. After transmission of the transgene to the F1, the β-globin locus was checked for

integrity using southern blots with the LCRε and γγδβ cosmid probes and specific probes within the LCR and the different genes. Transgene copy numbers were determined using as probes the βIVS2 fragment and a 0.9 kb Pvu I fragment from the endogenous mouse carbonic anhydrase II (CA-II) gene. The ratios of intensities of the βIVS2 /CA-II bands obtained for the PAC transgenics were compared to those obtained for the single copy PAC human β-globin locus transgenic lines [233]. Analysis was performed by PhosphorImager using ImageQuant software (Molecular Dynamics).

S1 Nuclease protection

S1 nuclease protection analysis was carried out with total RNA from 12.5dpc and 14.5dpc fetal livers. RNA was isolated using the Trizol reagent according to the manufacturer's instructions (Invitrogen). Conditions for S1 nuclease protection assays and polyacrylamide gel electrophoresis were essentially as previously described [7, 28, 199, 246]. The probes used were as follows: Human probes; γ-globin 5' probe, 320 bp *Ava*II fragment, protected fragment size 165 bp; β-globin 5' probe, 525 bp Acc I fragment, protected fragment size 155 bp. Mouse probe: α-globin, 300 bp *Bam*HI fragment, protected fragment size 185 bp bp. Quantitation was done on a PhosphorImager using the ImageQuant software (Molecular Dynamics).

Nuclear extracts and gel mobility shift assay

Extracts were prepared as described previously [173, 247]. A sense Lac^O oligomer (50 ng) was radioactively labeled for 30 minutes at 37°C by T4 polynucleotide kinase (Invitrogen) according to manufacturer's protocol. Adding 10 mM Tris pH7.5; 1 mM EDTA; 0.5% SDS, stopped the reaction and the labeled oligomer was purified over a Sephadex G50 column. Next, the labeled sense oligomer was annealed to a 5-fold excess of unlabeled antisense oligomer in 10 mM Tris pH7.5; 50 mM NaCl by heating for 10 minutes at 95°C followed by a slow cool down period. For gel mobility shift, 0.5 ng of labeled double-stranded Lac^O oligomer was incubated with 4 μg of nuclear extracts for 20 minutes at room temperature in 10 mM Tris pH7.5; 10 mM $MgCl_2$; 2 mg/ml poly(dIdC). Additionally, mouse α-GFP antibodies (Roche), 10x, and 100x excess of unlabeled double-stranded Lac^O oligomer was added. Samples were run on a 4% polyacrylamide gel at 300 Volts (4°C) and imaging was performed on a PhosphorImager using the ImageQuant software (Molecular Dynamics). Oligomers: sense 5'-CAC ATG TGG AAT TGT GAG CGG ATA ACA ATT-3'; antisense 5'-AAT TGT TAT CCG CTC ACA ATT CCA CAT GTG-3'.

Acknowledgements

We thank Gert van Cappelen and Adriaan Houtsmuller for technical assistance regarding confocal microscopy. Mariken de Krom and Ali Imam for advice on PAC construct related work. Ernie de Boer and Mariëtte van de Corput for advice and sharing data. John Kong-a-San, Ton de Wit, Rien van Haperen, Dubravka Drabek, and Albert Wai for micro-injection off the various constructs.

Chapter 3

Looping and interaction between hypersensitive sites in the active β-globin locus

Looping and interaction between hypersensitive sites in the active β-globin locus

BasTolhuis[1], Robert-Jan Palstra[1], Erik Splinter, Frank Grosveld and Wouter de Laat[2]

Department of Cell Biology and Genetics, Erasmus MC, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

[1] These authors contributed equally
[2] Corresponding author: w.delaat@erasmusmc.nl

Modified from: ***Molecular Cell*, 2002, Vol. 10, No. 6, 1454-1465**

Summary

Eukaryotic transcription can be regulated over tens or even hundreds of kilobases. We show that such long-range gene regulation *in vivo* involves spatial interactions between transcriptional elements, with intervening chromatin looping out. The spatial organisation of a 200 kb region spanning the murine β-globin locus was analysed in expressing erythroid and non-expressing brain tissue. In brain, the globin cluster adopts a seemingly linear conformation. In erythroid cells the hypersensitive sites of the Locus Control Region (LCR), located 40-60kb away from the active genes, come in close spatial proximity with these genes. The intervening chromatin with inactive globin genes loops out. Moreover, two distant hypersensitive regions participate in these interactions. We propose that clustering of regulatory elements is key to creating and maintaining active chromatin domains and regulating transcription.

Introduction

Transcriptional activation in higher eukaryotes frequently involves the long-range action of a number of regulatory DNA elements. Although this has been recognised for more than 20 years, it is still not clear how enhancers [248, 249], LCRs [16] or insulators/boundaries [29, 31, 225, 250] exert their effect on the process of chromatin modification and transcription over distance (up to hundreds of kilobases). Many different models have been put forward to explain distant effects. The 'looping model' states that enhancers and promoters communicate through direct interactions between proteins bound to the DNA elements, with the intervening DNA looping out. [213, 219, 251]. Other models imply a role for the DNA in between to support the transmission of some signal from enhancer to promoter. Direct support for the latter type of models comes from bacteria. Here, activation of the phage T4 late genes was found to involve loading on and sliding from the enhancer of trimeric gp45 along the DNA to the promoter to allow the forming of the transcription initiation complex [224]. The 'looping model' also receives support from studies on transcriptional regulation of many different prokaryotic genes. In fact, the model was originally based on work on bacterial and phage repressor proteins, like the Gal-, AraC and λ repressor proteins, which were found to function only when homo-multimerized and bound to two separate operator sites. Electron microscopy visually demonstrated the DNA in between to loop out (reviewed in [219]). Thus, both type of mechanisms appear to function in bacteria. Eukaryotes have more complex gene clusters with regulatory elements functioning over much greater distances. To date, there are no data that unambiguously demonstrate one (or more or combinations) of the models to be correct for the regulation of a given eukaryotic locus. Support for models has come from indirect and/or *in vitro* observations and often the distinction between the activation and actual transcription of a locus is not made. However with respect to transcription, a number of observations can only easily be explained by the 'looping model'. The first type of experiments involves studies on *trans*-activation, i.e. the ability of an enhancer to activate a promoter present on a physically separate DNA molecule. Most important in this respect is the naturally occurring phenomenon of transvection in *Drosophila* [252]. In addition, Schaffner and co-workers demonstrated *in vitro* that enhancers can stimulate transcription *in trans*, by coupling an enhancer- to a promoter-containing plasmid via a biotin-streptavidin bridge [253]. Similarly, *trans*-activation of transcription was observed when enhancer-containing and promoter-containing plasmids were injected as intertwined catenates into frog oocytes [254]. More recently, transient transfection assays with reporter plasmids and GAGA as a DNA-bridging factor also demonstrated transcriptional activation *in trans* in mammalian cells [255]. All these studies on *trans*-activation demonstrate that a cis configuration of enhancer and promoter is not an absolute prerequisite for interaction, as predicted only by the 'looping model'.

In addition, gene competition for a single regulator [9, 213, 248], leading to alternate

transcription [111, 216, 217], is also most easily explained by 'looping', particularly because the competitive advantage of the enhancer-proximal gene is lost when the genes are closely spaced at further distance from the regulator [213, 214, 256]. Finally, in yeast, a downstream enhancer was recently demonstrated to activate gene expression from a distance, by making use of loops induced by telomeres [257]. However all these experiments were either done *in vitro* or are indirect in nature. None of them directly shows *in vivo* that two distal elements linked *in cis* interact by coming in close spatial proximity with intervening DNA looping out.

Here, we provide evidence that looping occurs during transcription *in vivo*. We demonstrate that the murine β-globin LCR is in physical proximity to the active globin genes *in vivo* in expressing tissue with the intervening DNA looping out. Interaction and looping are not observed in non-expressing tissue. In addition, DNase I hypersensitive sites at both end of the locus participate in these interactions, again by looping out intervening DNA. Thus, multiple hypersensitive sites spread over 130 kilobases interact to form a cluster in the nuclear space. On the basis of these data we propose that direct interactions between distal DNase I hypersensitive sites and looping out of chromatin is crucial in establishing an open chromatin domain and activating transcription.

Results

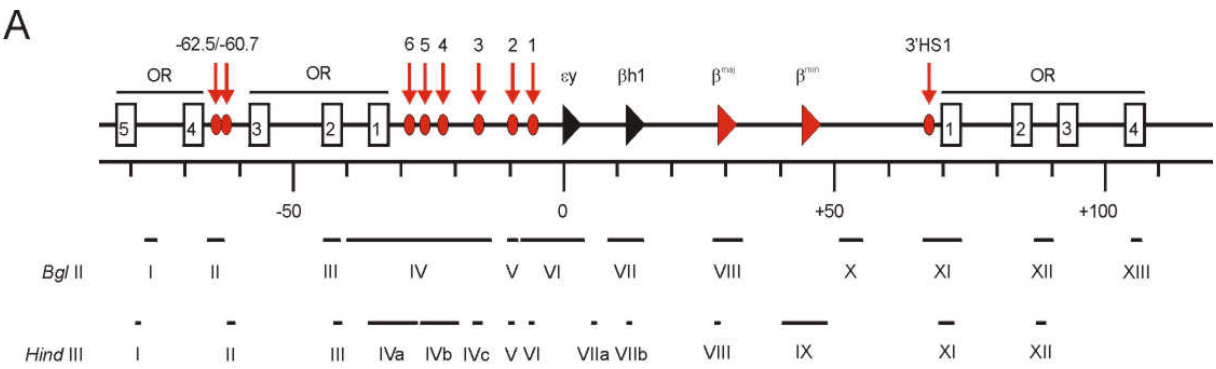
Applying 3C technology to the murine β-globin locus

We applied methodology recently developed by Dekker et al. [229] to gain insight into long-range interactions between the LCR and the genes in the murine β-globin locus. The principle of this technique, Chromosome Conformation Capture (3C), is that cells are treated with formaldehyde to cross-link proteins to other proteins nearby and DNA (see also figure 3.1B). The resulting DNA-protein network is then subjected to cleavage by a restriction enzyme, which is followed by ligation at low DNA concentration. Under such conditions, ligations between cross-linked DNA fragments, which is intramolecular, is strongly favored over ligations between random fragments, which is intermolecular [229]. After ligation, the cross-links are reversed and ligation products are detected and quantified by polymerase chain reaction (PCR). The cross-linking frequency of two specific restriction fragments, as measured by the amount of corresponding ligation product, is proportional to the frequency with which these two genomic sites interact [229]. Thus 3C analysis provides information about the spatial organisation of chromosomal regions *in vivo*.

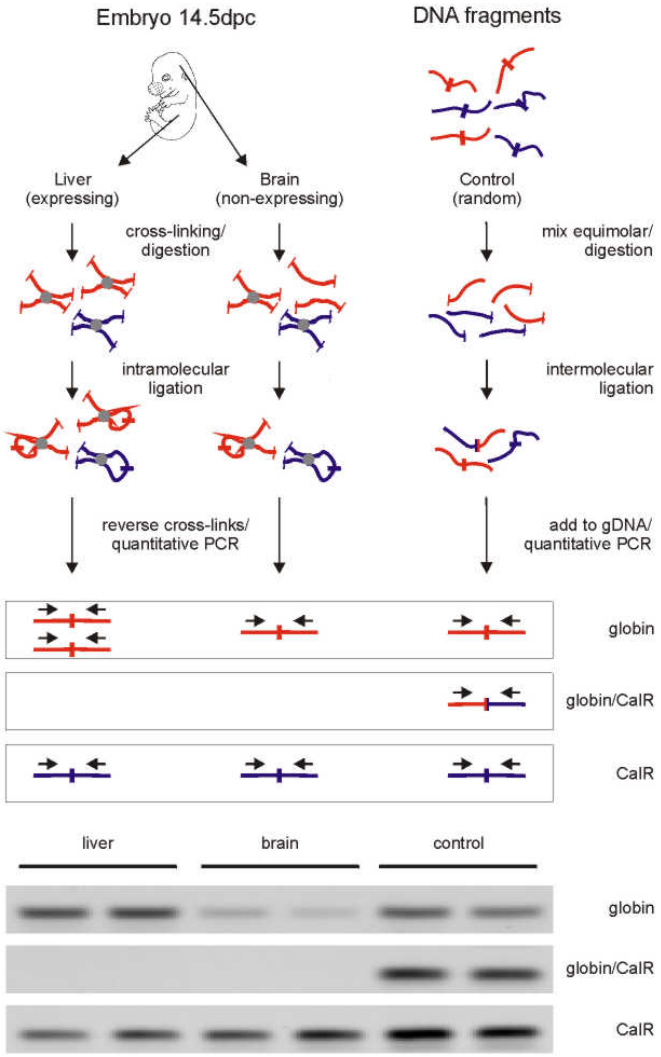
Figure 3.1. 3C technology in the murine β-globin locus

(A) Schematic presentation of the murine β-globin locus. Red arrows and ellipses depict the individual HSs. The globin genes are indicated by triangles, with active genes (β^{maj} and β^{min}) in red, and inactive genes (εy and βh1) in black. The white boxes indicate the olfactory receptor (OR) genes (5'OR1-5 and 3'OR1-4). The two sets of restriction fragments (*Bgl*II and *Hind*III) that were used for 3C-analysis are shown below the locus. The individual fragments are indicated by roman numbers. Identical numbering between *Bgl*II and *Hind*III indicates that two fragments colocalize. Distances (roman numerals) are in kb counting from the site of initiation of the εy gene.
(B) Schematic outline of the 3C-analysis. Globin fragments (red), CalR fragments (blue), restriction sites (perpendicular bars on fragments), cross-links and PCR primers are indicated. Examples of PCR results (always done in duplo) show products obtained with HindIII globin fragments VIII and IV-b (top), globin fragment VIII and one of the HindIII CalR fragments (middle) and the two HindIII CalR fragments (bottom). Tissue lanes in middle panel were always empty, with every globin fragment tested. The CalR products (bottom) were used for normalizing signals.

A



B



C

$$X(gl) = \frac{[A(gl) / A(CalR)]_{tissue}}{[A(gl) / A(CalR)]_{control}}$$

A schematic presentation of the murine β-globin locus is given in figure 3.1A. Briefly, the locus contains an LCR, comprising 6 HSs (5'HS1-6), two embryonic genes, εy and βh1 (expressed in the yolk sac), and two adult genes, βmajor and βminor (expressed in fetal liver and adult spleen/bone marrow). The LCR is required for high levels of expression of all β-globin genes. Similar to what is observed in the human β globin locus, the murine β-globin locus is flanked by olfactory receptor (OR) genes, which are inactive in globin-expressing erythroid tissue [161, 162]. Also similar to the human locus are the strong erythroid-specific DNase I HS at the 3' side (3'HS1) between βminor and the OR genes and two closely spaced HSs (HS –60.7 and HS-62.5) at the far 5' side located between 5'OR3 and 5'OR4 [164].

Two independent sets of restriction fragments (BglII- and HindIII-fragments, respectively) were used for 3C-analysis of the β-globin locus. Each set covers the 200 kb region depicted in figure 3.1A, with intervals between analysed DNA fragments of approximately 20 kilobases or smaller. Analysis was performed on 14.5 dpc mouse fetal livers, which express the most distal globin genes, βmajor and βminor. Brain from the same 14.5 dpc embryos was simultaneously analysed as a non-expressing control tissue.

A number of experimental controls were included. Firstly, we checked the efficiency of restriction enzyme digestion. Southern blotting and PCR analysis showed that the restriction sites analysed were cleaved without any preference for any particular region(s) after overnight incubation with an excess of enzyme (data not shown). Secondly, we determined the range of amount of template that shows linear PCR product formation. Similar ranges were found with both liver and brain template (data not shown), and roughly equal amounts (~300 ng DNA template per reaction) were used in all subsequent experiments. Thirdly, to correctly interpret signal intensities obtained with a given primer set by quantitative PCR, one needs to correct for the PCR amplification efficiency of that set. Thus, a control template is required in which all possible ligation products are present in equimolar amounts. In yeast, this was done by digesting and randomly ligating non-cross-linked genomic DNA [229]. For mammalian cells, with a genome one hundred times the size of the yeast genome, we found that random ligation of two specific loci is too rare an event to be detected by PCR. We therefore enriched for ligation products of interest by mixing equimolar amounts of DNA fragments that span each of the restriction sites analysed (see figure 3.1B). After digestion and ligation, this mix was added to genomic DNA to serve as a control template (see also Experimental Procedures). As a result, the cross-linking frequency between two loci can be expressed as the ratio of signal obtained by quantitative PCR on cross-linked template versus that obtained on control template. Fourthly, we measured the cross-linking and ligation efficiencies in both tissues to be able to compare cross-linking frequencies. This was done by comparing the cross-linking frequency between two restriction fragments present on an unrelated locus situated on another chromosome. Two neighbouring fragments were used, with the restriction sites analysed ~1.5 kilobases apart, in the transcribed part of the calreticulin locus (CalR) on chromosome 8 (the β-globin locus is on mouse chromosome 7). The CalR locus, embedded in an area of ubiquitously expressed genes, is expressed at similar levels in 14.5 dpc brain and liver (WdL, unpublished results). It is therefore reasonable to assume that it adopts a similar spatial conformation in both tissues. Thus by normalizing each cross-linking frequency to the cross-linking frequency observed between the CalR-fragments within a tissue, we could correct for differences in amount and quality of template. Similarly, by normalizing the observed random ligation efficiency of two given fragments to that observed of the CalR fragments, we corrected for differences in the amount of control template between experiments. The equation used to calculate the relative cross-linking frequency is given in figure 3.1C. As a result of this normalisation, the "cross-linking frequency" value 1 arbitrarily corresponds to the cross-linking frequency between our control CalR fragments. Finally, the cross-linking frequencies between globin fragments and CalR fragments were always measured as an additional control. As expected for the interaction between two unrelated loci,

globin-CalR cross-linking frequencies were always found to be zero (no PCR signals observed in tissues, see figure 3.1B).

The globin locus adopts a linear conformation in non-expressing brain cells

We performed 3C-analysis on expressing and non-expressing tissue from 14.5 dpc embryos to be able to relate the spatial conformation of the β-globin locus to its transcriptional status. Figure 3.2 shows results obtained in a non-expressing tissue, the brain. Depicted are locus-wide cross-linking frequencies for two different BglII fragments ('fixed' fragments), one in the middle of the 200 kb region (fragment V) and one at the 5' end (fragment II). The central fragment V, a relatively small fragment containing HS2 of the LCR, showed the highest cross-linking frequency with the closest fragments IV and VI. Cross-linking frequency gradually decreased with fragments located further away on the linear DNA template (figure 3.2A). No significant peaks of interactions were observed between fragment V and more distal DNA fragments. Similar results were obtained for the DNA fragment at the 5' end of the region (II) (figure 3.2B). Thus in brain, we observed a direct correlation between spatial proximity and distance along the linear β-globin DNA template. This holds for any 'fixed' fragment in this region, independent of the restriction enzyme used (see data below). Such a correlation between distance in space and distance in kilobases would be expected of a linear structure [228]. Hence we conclude that the 200 kb region encompassing the β-globin locus adopts an essentially linear conformation in the nucleus of the non-expressing brain cell.

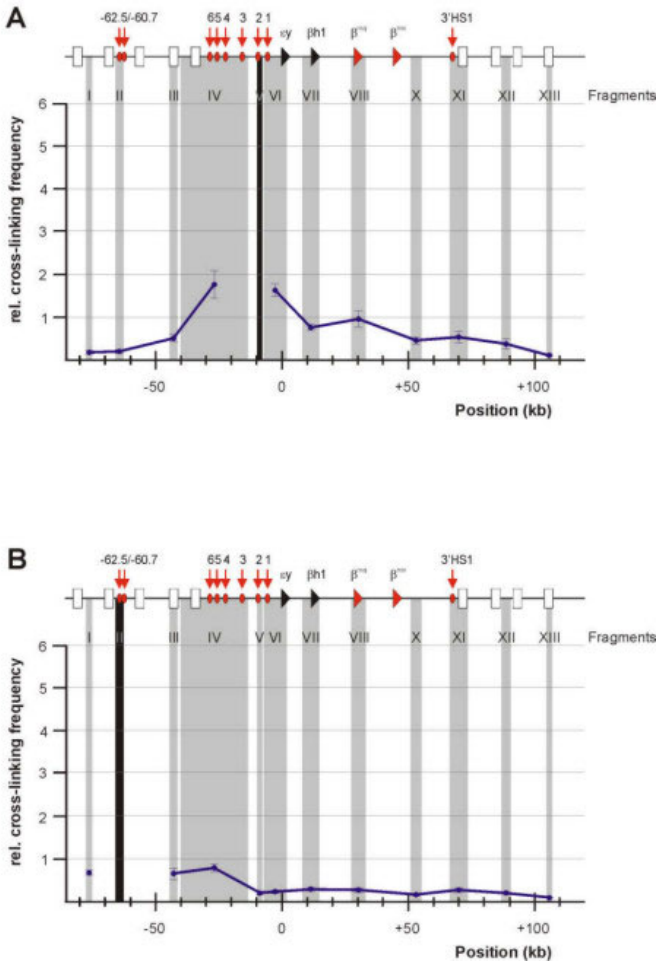


Figure 3.2. Linear conformation of the β-globin locus in non-expressing brain cells.

The murine β-globin locus is depicted on top of each graph (for explanation of symbols, see figure 1A). X-axis shows position in the locus. Black shading shows the position and size of the 'fixed' fragment. Grey shading indicates position and size of other fragments. Standard-error-of-mean is indicated. Cross-linking frequency with a value of 1 arbitrarily corresponds to the cross-linking frequency between two neighbouring CalR control fragments (with restriction sites analysed being 1.5 kb apart). Scaling on Y-axis (from 0 to 6) allows direct comparison with Figures 3-6. (A). Relative cross-linking frequencies between 'fixed' *Bgl*II fragment V (5'HS2 in LCR) and the rest of the locus. (B) Relative cross-linking frequencies between 'fixed' *Bgl*II fragment II (5'HS-62.5/60.7) and the rest of the locus.

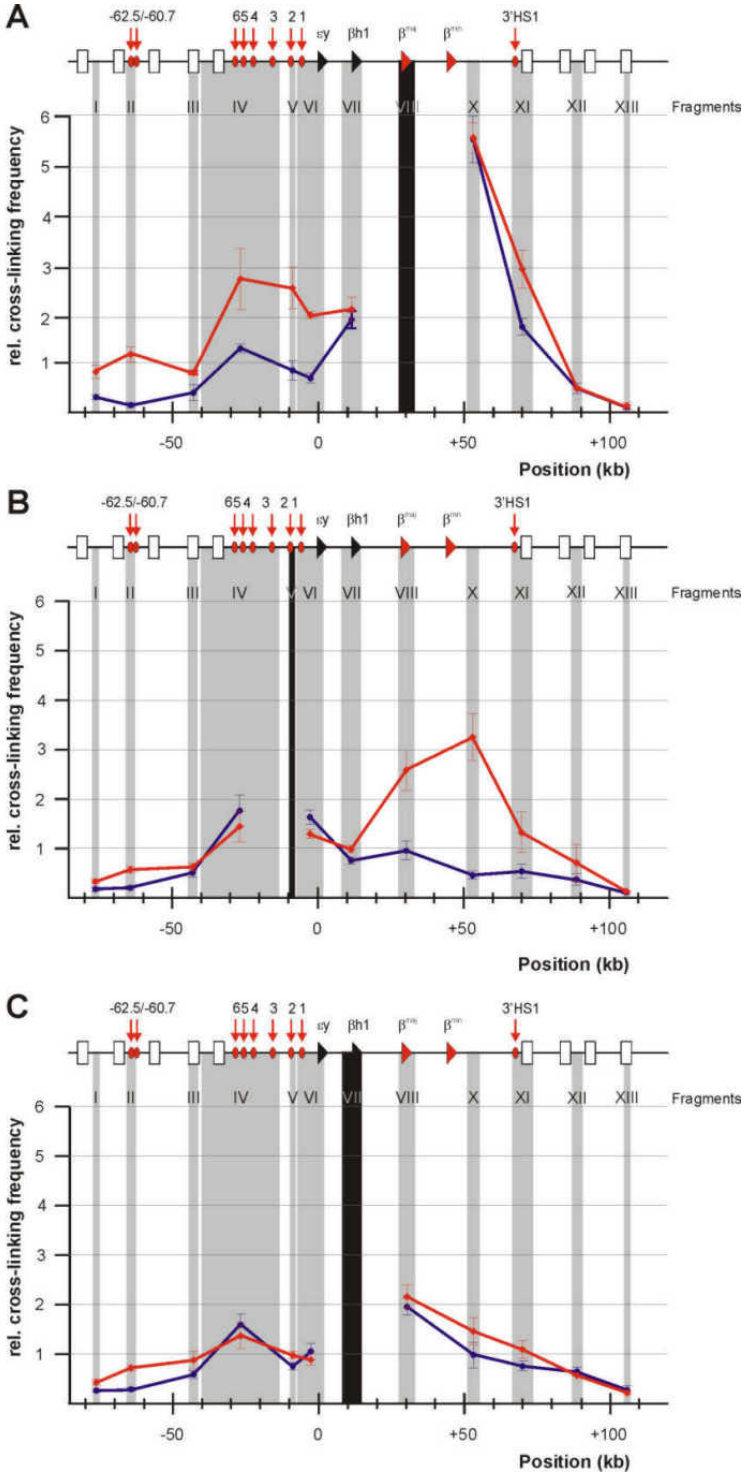


Figure 3.3. Erythroid-specific interaction and looping between the LCR and an active β-globin gene.

Relative cross-linking frequencies observed in fetal liver are shown in red. For comparison, data obtained in brain are depicted in blue. Standard-error-of-mean is indicated. Cross-linking frequency with a value of 1 arbitrarily corresponds to the cross-linking frequency between two neighbouring CalR control fragments (with restriction sites analysed being 1.5 kb apart). Scaling on Y-axis (from 0 to 6) allows direct comparison with Figures 2 and 4-6. (A) 'Fixed' *Bgl*II fragment VIII (β^{maj}) versus the rest of the locus. (B) 'Fixed' *Bgl*II fragment V (5'HS2) versus the rest of the locus. (C) 'Fixed' *Bgl*II fragment VII (βh1) versus the rest of the locus.

Spatial interaction and looping between the LCR and the active genes in the expressing fetal liver

Next, we analysed the spatial organisation of the β-globin locus in the expressing 14.5 dpc fetal liver cells. The active globin genes, βmajor and βminor, are 34 and 49 kb away from the 3' side of the LCR, respectively. We first focussed on a BglII fragment (fragment VIII) containing the active βmajor gene with all the known local regulatory elements, including the promoter and the enhancer ~1 kb downstream of the transcribed sequence. In agreement with the findings presented above, the curve

for brain was indicative of a linear conformation (figure 3.3A). In fetal liver, cross-linking frequencies identical to those in brain were observed for fragments closest to fragment VIII. However, when DNA elements more towards the 5' side of the region were analysed, up to 3-fold elevated cross-linking frequencies were found in liver as compared to brain with fragments IV, V and VI. Most interestingly, these are the three BglII fragments that together cover all six hypersensitive sites of the LCR. Beyond the LCR, even further 5' from the βmajor gene, cross-linking frequencies dropped again to the levels observed in brain (with the exception of fragment II, discussed below). These data indicate that in the nucleus of the expressing fetal liver cell, the active βmajor gene comes in close vicinity to the LCR.

This is confirmed when the reciprocal experiment is carried out using an LCR fragment as the fixed fragment. BglII restriction sites flank HS2 of the LCR, resulting in fragment V. When this fragment was tested versus the others in fetal liver, fragment VIII (βmajor), but also fragment X, containing DNA sequences just 3' of the active βminor gene, showed highly elevated cross-linking frequencies in fetal liver compared to brain (figure 3.3B). In fact, in fetal liver, but not in brain, the cross-linking frequency between HS2 and the active adult genes is much higher than that between HS2 and the inactive, embryonic genes (εy and βh1, present on fragment VI and VII, respectively). Thus, these data show that in expressing cells, the β-globin LCR and the distal active genes come in physical proximity, whereas the inactive genes appear to be located further away from the LCR fragment V.

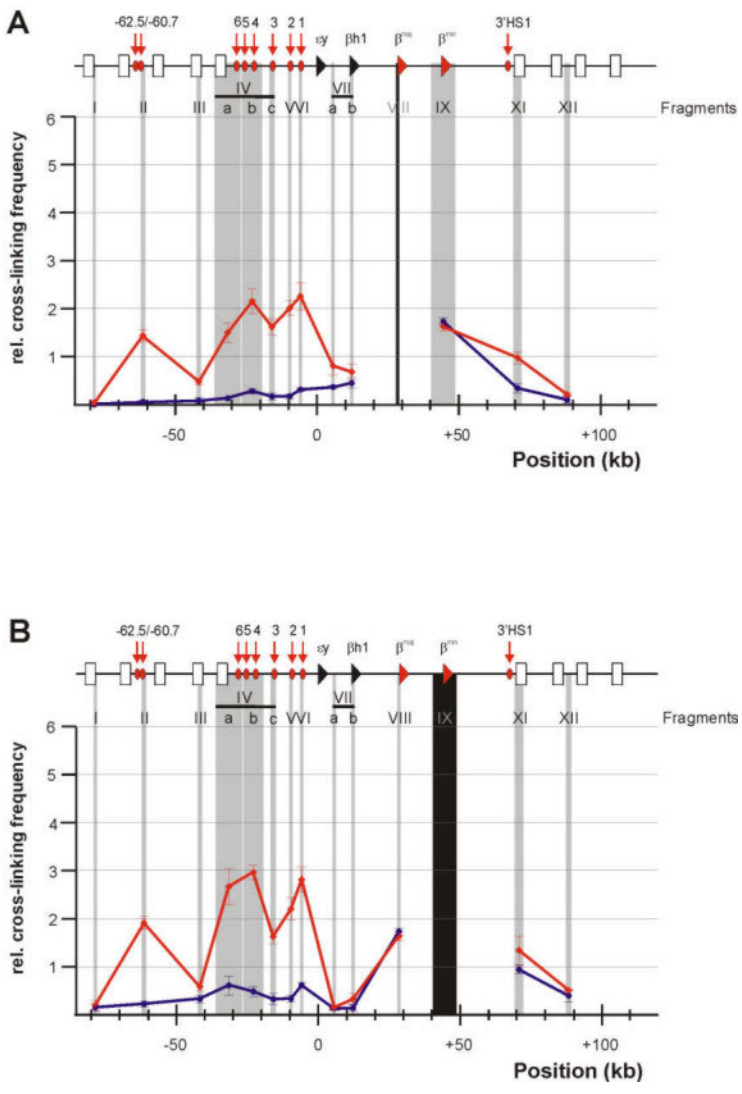


Figure 3.4. Erythroid-specific interactions between the active β-globin genes and individual hypersensitive sites in the LCR.

Relative cross-linking frequencies observed in fetal liver (red) and brain (blue) are shown. Standard-error-of-mean is indicated. Cross-linking frequency with a value of 1 arbitrarily corresponds to the cross-linking frequency between two neighbouring CalR control fragments (with restriction sites analysed being 1.5 kb apart). Scaling on Y-axis (from 0 to 6) allows direct comparison with other figures. (A) 'Fixed' *Hind*III fragment VIII (β^{maj}) versus the rest of the locus. (B) 'Fixed' *Hind*III fragment IX (β^{min}) versus the rest of the locus.

In order to determine whether in fetal liver the inactive genes indeed do not come in close proximity to other sequences in the locus, we looked at locus-wide cross-linking frequencies of the βh1 gene (fragment VII). Almost identical cross-linking frequencies between βh1 and the rest of the locus were observed in liver and in brain for both a BglII (figure 3.3C) and HindIII digest (not shown). Similar results were obtained for a HindIII fragment close to εy (VII-a, not shown & see Fig. 4-6). This suggests that the inactive genes are not interacting with the LCR. We conclude that the LCR interacts specifically with the active distal β-globin genes with intervening DNA containing the inactive genes looping out.

All hypersensitive sites of the LCR participate in the long-range interactions

Whereas BglII cuts relatively infrequently in the murine β-globin locus, resulting in the large fragments analysed and described above, digestion by HindIII yields smaller DNA fragments, which may allow fine-mapping of the interactions. Most relevant to our studies, HindIII cuts in between most of the hypersensitive sites of the LCR (with the exception of HS4 and HS5, which are present on one HindIII fragment). Analysis of cross-linking frequencies with a fixed HindIII fragment VIII, containing 300 base-

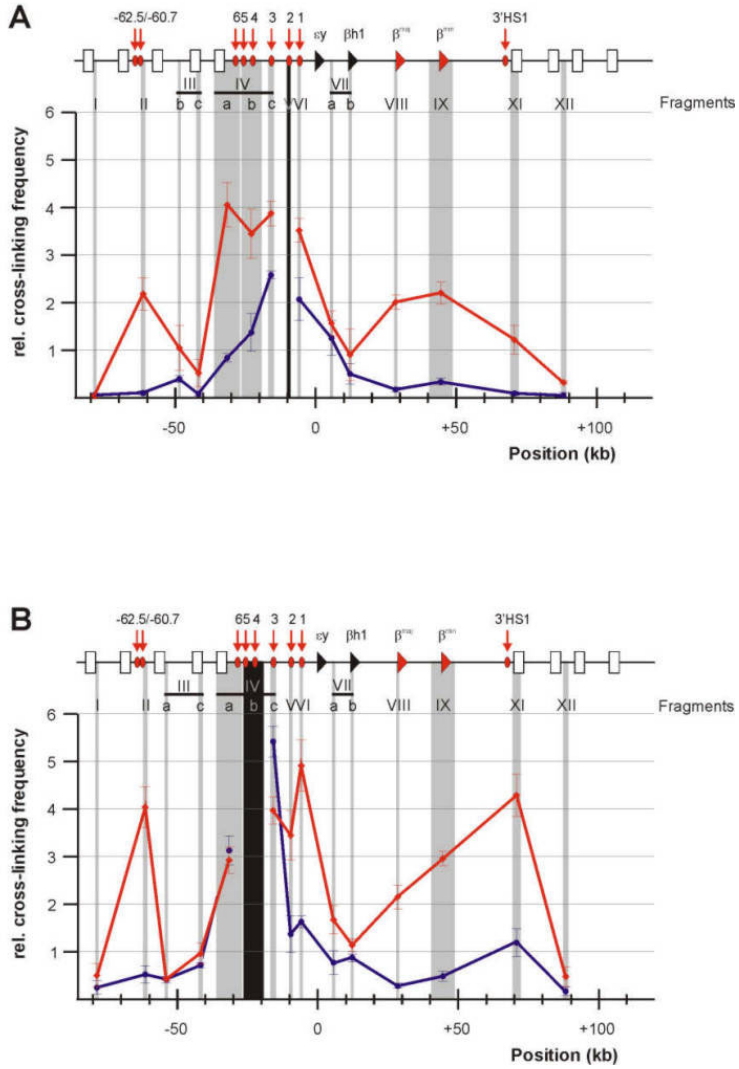


Figure 3.5. Erythroid-specific high cross-linking frequencies among the individual hypersensitive sites of the LCR and two distal hypersensitive sites.

Relative cross-linking frequencies observed in fetal liver (red) and brain (blue) are shown. Standard-error-of-mean is indicated. Cross-linking frequency with a value of 1 arbitrarily corresponds to the cross-linking frequency between two neighbouring CalR control fragments (with restriction sites analysed being 1.5 kb apart). Scaling on Y-axis (from 0 to 6) allows direct comparison with other figures. (A) 'Fixed' *Hind*III fragment V (5'HS2 of the LCR) versus the rest of the locus. (B) 'Fixed' *Hind*III fragment IV-b (5'HS4-5 of the LCR) versus the rest of the locus

pairs of the βmajor promoter plus one third of the coding part of this gene, confirmed the fetal liver-specific interaction with the LCR (figure 3.4A). In fact, elevated cross-linking frequencies with the βmajor fragment were observed for all fragments containing a hypersensitive site of the LCR (fragments IV-a, -b and -c, and fragment V and VI). As seen in the BglII experiments, cross-linking frequencies with βmajor dropped for fragments flanking the LCR (again with the exception of fragment II, discussed below). Thus, the HindIII data indicate that all individual hypersensitive sites of the LCR (HS1-6) participate in long-range interaction. The same results were obtained with fragment IX, encompassing the active βminor gene (figure 3.4B), although here the data suggest that HS2 (fragment V) and HS3 (fragment IV-c) do not participate as actively in the interaction as the other hypersensitive sites do. This may indeed be the case, but it may equally well reflect a technical problem (see discussion). Nevertheless these data strongly support the hypothesis that the individual hypersensitive sites of the LCR act together to contact distal genes in the fetal liver [15, 111].

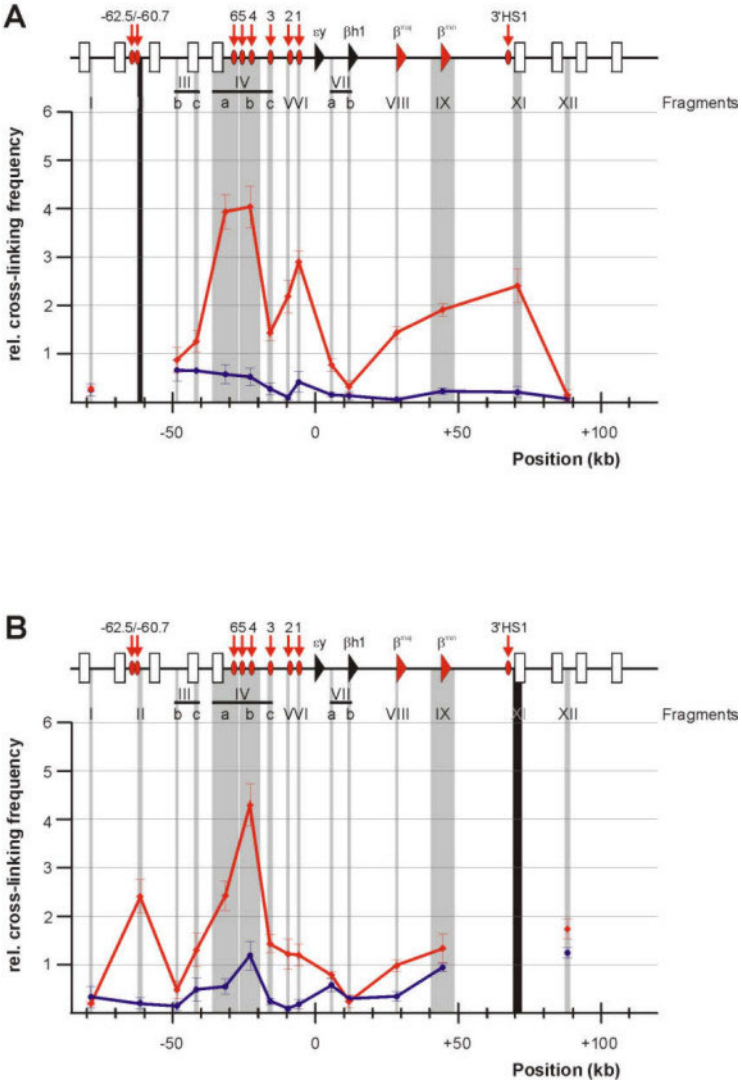


Figure 3.6. Two distal hypersensitive sites at each side of the locus cluster with the LCR and the genes.

Relative cross-linking frequencies observed in fetal liver (red) and brain (blue) are shown. Standard-error-of-mean is indicated. Cross-linking frequency with a value of 1 arbitrarily corresponds to the cross-linking frequency between two neighbouring CalR control fragments (with restriction sites analysed being 1.5 kb apart). Scaling on Y-axis (from 0 to 6) allows direct comparison with other figures. (A) 'Fixed' *Hind*III fragment II (5'HS-62.5/-60.7) versus the rest of the locus. (B) 'Fixed' *Hind*III fragment XI (3'HS1) versus the rest of the locus.

If indeed the LCR forms one spatial entity in expressing cells, tissue-specific high cross-linking frequencies among the individual hypersensitive sites of the LCR would be expected. This is indeed what we observe. For example, taking HS2 (fragment V) as the 'fixed' fragment, we found fetal liver-specific high cross-linking frequencies with all other hypersensitive sites of the LCR (figure 3.5A). Similar results were obtained with fixed fragment IV-b (HS4-5, figure 3.5B), IV-a, IV-c, and VI (HS6, HS3 and HS1 respectively, data not shown). Together, these data provide strong support for the LCR acting as a 'holocomplex' in erythroid cells to activate the globin genes.

HSs at both ends of the locus participate in the interactions between the LCR and the active genes

Two other erythroid-specific interactions stand out. In figure 3.5B, for example, high cross-linking frequencies were observed between HS4/5 and the fragments II and XI, at the far 5' and 3' end of the region, respectively. Interestingly, fragment II contains (part of) the recently identified hypersensitive sites –62.5 and –60.7 [164], and fragment XI is located just 3' of another erythroid-specific hypersensitive site, 3'HS1 [16, 163]. Interaction with both of the distal hypersensitive sites was seen with all other hypersensitive sites of the LCR, both in the HindIII experiments (see figure 3.5A and data not shown) and in the BglII experiments (see figure 3.3B, and data not shown). Moreover the active βmajor and βminor genes also showed erythroid-specific interactions with 5'HS62.5/-60.7 (figures 3.3A, 3.4A and 3.4B), despite being approximately 100kb away. These data suggest a complex series of interactions between hypersensitive sites in the β-globin locus in expressing tissue.

To further investigate this, we analysed locus-wide interactions with the distal hypersensitive sites. Figure 3.6A shows the results for fragment II, which confirm the interaction between 5'HS-62.5/-60.7 and LCR elements in the fetal liver. Fragments I and III, flanking these 5'HS, do not participate in this interaction (both in the BglII and HindIII digestions), suggesting that the intervening DNA loops out. High cross-linking frequencies were also found between 5'HS-62.5/-60.7 and 3'HS1, which is remarkable considering the two sites are 130 kb apart on the linear chromatin template. Comparable interactions were observed using 3'HS1 as the 'fixed' fragment (figure 3.6B). However it should be noted that the data for 3'HS1 are similar to those found for βmajor and βminor and that this region appears to act as one block. The latter may point at some compaction, perhaps caused by the large amount of repetitive DNA present in this region [161]. Nevertheless, our data demonstrate that all the hypersensitive sites and the active genes of the β-globin locus cluster together in space in the erythroid nucleus.

Discussion

The 'looping model' postulates that regulatory elements and genes/promoters communicate through direct interactions between proteins bound to the DNA, with intervening chromatin looping out. In this paper we have demonstrated that the distal regulatory elements and the active genes, which are linked *in cis* in the murine β-globin locus, interact *in vivo* while the intervening DNA loops out. This looping is only seen in expressing cells and provides direct *in vivo* evidence for the 'looping model'. Previous support for this model has come from several types of studies. *Trans*-activation, i.e. the ability of an enhancer to activate a promoter located on a physically separate DNA molecule, is most easily explained by direct contact between the enhancer and the gene. This has been observed in transvection in *Drosophila* [252] and in a number of *in vitro* experiments with artificial DNA constructs [253-255]. Competition between genes for a single regulator [9, 213, 248] leading to alternate transcription [111] is also most easily explained by looping, particularly because the competitive advantage of the enhancer proximal gene is lost when the genes are closely spaced at distances further from the enhancer [214, 256]. However, all this evidence is indirect and each can also be

explained by other mechanisms. The findings presented here show direct evidence for looping in the active β-globin locus, whereas a linear type of structure is found for the non-expressing locus. In particular, the observation that two hypersensitive sites at the far ends of the region cluster with the LCR and the active genes (i.e. all hypersensitive sites) provides new insights into long-range interactions (see below). However, the limitations of the 3C technique should also be noted in order to avoid overinterpretation of the results.

Interpreting 3C-analysis of the β-globin locus

Some technical and biological aspects of the results by 3C-analysis should be considered. As pointed out originally by Dekker et al. [229], measuring cross-linking efficiency by the formation of ligation products largely depends on the frequency with which two genomic sites interact. They showed that contributions of other parameters, such as local protein concentrations or a favorable geometry of the cross-linked intermediate, are minor. Our results support this notion. However, we further believe that additional parameters, e.g. the fragment size, notably affect the cross-linking efficiency. Comparison of cross-linking frequencies observed with the large (26kb) BglII fragment IV (covering HS3-6 of the LCR and 12 kb upstream), to those observed with the much smaller HindIII fragments IV-a, -b and –c (containing HS6, 4-5 and 3 as separate entities) reveals an increased background in brain for the large fragment. This can be explained by assuming that the chance of being crosslinked per se increases with fragment size. Also, an increase in ligation to irrelevant fragments will compete with ligation to specifically interacting fragments, causing underestimation of specific interactions in the fetal liver. Thus, to determine whether a specific interaction occurs between two given DNA sequences, it is best to study smaller fragments containing isolated entities.

The accuracy of signals obtained with the control template is crucial for our analysis. Since cross-linking values in brain and in liver are both normalized to the same control value, we were concerned about the fact that HindIII fragment IV-c showed a dip in relative cross-linking frequency with every fragment tested, both in brain and in liver. This result was due to high PCR signals in the control rather than low signals in the tissue samples (data not shown). Designing new primers did not solve this problem. Thus, although the observed cross-linking frequencies with HindIII fragment IV-c may be real, it is more likely that it reflects an as yet unresolved technical issue.

Purely biological parameters also play a role. For example, in 14.5 dpc fetal liver about 15-20% of the cells is not expressing globin (judged by many RNA FISH experiments). These are likely to adopt a conformation similar to that observed in brain and contribute to the total amount of substrate in the ligation reaction, but not to the specific ligation frequency. Thus the real value of erythroid-specific interactions will be underestimated, which increases the significance of finding these interactions, particularly the ones over large distances. Perhaps most importantly, interactions between distal DNA elements are thought to be dynamic [111], while these measurements represent steady-state average levels. For example, a very important, but short-lived interaction for transcription initiation [111] may score much lower than a more long-lived interaction that would only stabilise the complex.

Given these limitations, and the unknown dimensions of the chromatin fiber in the globin locus *in vivo*, the results presented here do not allow a strictly quantitative interpretation or conclusions as to what HS is responsible for a given interaction and/or function. Predictions about the dynamics of the interactions or real nuclear distances are therefore not possible at this stage of development of the technique.

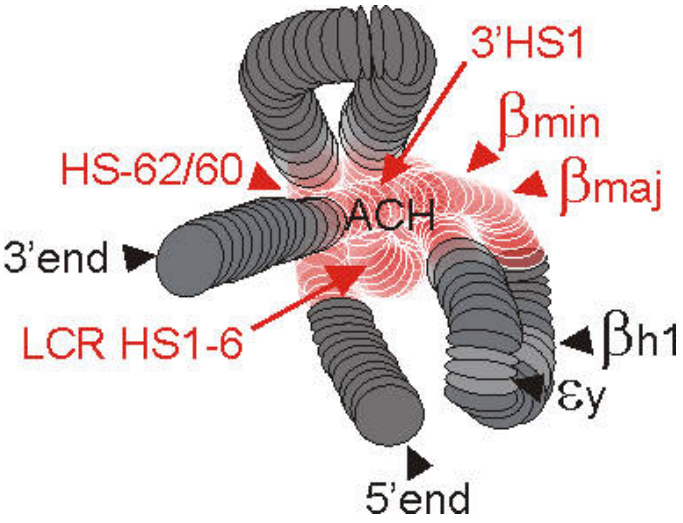


Figure 3.7. A 3D model of the ACH.

A hypothetical model of the Active Chromatin Hub (ACH) is shown to illustrate the 3D nature of the ACH (not to scale), not the actual position of the elements relative to each other *in vivo*. Red indicates the active regions (hypersensitive sites and active genes) of the locus forming a hub of hyperaccessible chromatin (ACH). The inactive regions of the locus, having a more compact chromatin structure, are indicated in grey, with the inactive βh1 and εy genes in lighter grey. The olfactory genes are not shown. The interactions in the ACH would be dynamic in nature, in particular with the active genes (βmaj and βmin), which are alternately transcribed.

The hypersensitive sites, looping and an open chromatin domain

Despite the limitations of the 3C technique, we can conclude that the 6 hypersensitive sites of the LCR, HS1-6, interact with the active genes, βmajor and βminor in the 14.5 dpc fetal liver, with the inactive εy and βh1 genes on the intervening DNA fiber looping out. The upstream 5'HS-62.5/-60.7 participate in this interaction, again with the intervening DNA looping out. At the other end of the locus the 3'HS1 is also involved in the contacts, but we have no evidence for DNA looping out between the genes and 3'HS1. This region contains a large amount of repetitive DNA and may adopt a compacted structure as it appears to act in concert. The data also show a subdivision of the interactions, because we consistently observe the extreme 5' and 3' HS (5'HS-62.5/-60.7 and 3'HS1, respectively) to be closer to the 5' half of the LCR (HS4-6), which is not observed for the expressed genes.

The clustering of all hypersensitive sites in the β-globin locus is intriguing. Interactions are not confined to the outermost HSs (we cannot exclude the presence of even more distal erythroid-specific hypersensitive sites), as proposed in boundary models (for review, see [250]), nor to sequences that have been proposed to act as insulators [184], but include all HSs and the promoters/enhancers of the genes. Thus, rather than being a particular type of transcription element, hypersensitivity appears to be the determining criterion for a DNA element to participate in clustering. We anticipate that this clustering is not confined to the β-globin locus only. We propose to name a 3D clustering of hypersensitive sites an 'active chromatin hub' or ACH (figure 3.7). Its formation is required to initiate transcription in repressive chromatin surroundings. The affinity between distal DNA hypersensitive sites determines whether an ACH is productively formed or not. Affinity depends on the transcription factors bound to these DNA elements and can therefore be modulated [28, 181, 258]. Entry of new HSs may stabilize or destabilize existing interactions, which in turn can alter expression levels of genes present in the ACH. The model does not predict how DNA sequences become hypersensitive in the first place (e.g. by mass action [259]), but stabilisation/maintenance of hypersensitivity is proposed

to depend on ACH formation. Surrounded by less active chromatin, the ACH would create a biphasic system, ensuring and stabilising a high local concentration of transcription factors and associated chromatin modifying proteins to allow efficient transcription. The hypersensitive regions and promoters of the genes would have very high levels of for example histone acetylation [260, 261], whereas the chromatin outside the ACH would be less acetylated. An ACH need not occupy a fixed position in the nucleus, but can be a dynamic fluid entity, possibly inside the Interchromatin Domain (ICD) compartment [262]. We propose that stable formation of an ACH underlies position-independent expression in transgenic experiments, which indeed can be accomplished by various combinations of HSs. Such a scenario would explain why multicopy inserts may give position independent expression [15, 263].

Although formation of HSs in the LCR precedes transcription [168, 264], we presently do not know whether the same holds for ACH formation. However, it is tempting to speculate that the ACH would take shape first, creating the appropriate environment, by modification of the locus, to recruit the actual transcription machinery. The observation that the globin genes are alternately transcribed [111, 217] shows that only one of the genes is transcribed at any given moment. This implies that there is only one position of interaction within the ACH that allows initiation of globin gene transcription. In other gene clusters such a 'productive' interaction may become stabilised and explain for example single gene expression (Olfactory Receptor genes).

We presently do not know how looping in the β-globin locus is accomplished. Although we like to think that initial contact occurs through random collision between distal elements, we cannot exclude other mechanisms to be involved in loop formation. Also, we do not know whether sequences other than HSs (and cognate factors) participate directly in the ACH or perhaps stabilise its structure. Evidence from both Drosophila [265-269] and mammalian systems [157, 270, 271] strongly suggests that there are elements and protein factors that stabilise long range interactions. It will be interesting to determine whether such sequences are indeed part of the ACH.

Methods

Chromosome Conformation Capture (3C)

We used the procedure recently developed by Dekker and co-workers [229] with small adaptations to determine the spatial organization of the murine β-globin locus in 14.5 dpc embryos. Per experiment 10-12 fetal livers or fetal brains were resuspended in DMEM supplemented with 10% FCS. The equivalent of 2 fetal livers or 4 fetal brains (approximately $4x10^7$ cells) was diluted to 50 ml with DMEM (10% FCS). Formaldehyde was added to 2%, and the samples were cross-linked for 10 minutes at room temperature. The reaction was quenched by addition of glycine to 0.125M. Nuclei were harvested by lysis of the cells in ice-cold lysis buffer (10mM Tris, 10mM NaCl, 0.2% NP-40, pH=8.0) containing protease inhibitors. Nuclei were resuspended in the appropriate restriction buffer containing 0.3% SDS and incubated for 1 hour at 37°C while shaking. Triton X-100 was added to 1.8% and the nuclei were further incubated for 1 hour at 37°C to sequester the SDS. The cross-linked DNA was digested overnight with the restriction enzyme (BglII or HindIII). Overnight incubation at 37°C did not result in any specific loss of hypersensitive sites due to the action of endogenous nuclease activity (data not shown). The restriction enzyme was inactivated by addition of SDS to 1.6% and incubation at 65°C for 20 minutes. The reaction was diluted (to 2.5ng/μl of genomic DNA) with ligase buffer (30mM Tris-HCl, 10mM $MgCl_2$, 10mM DTT, 1mM ATP, pH 7.8) and Triton X-100 was added to 1% and incubated for 1 hour at 37°C. The DNA was ligated using T4 ligase for 4.5 hours at 16°C followed by 30 minutes at room temperature. Proteinase K was added and samples were incubated overnight at 65°C to reverse the cross-links. The following day samples were incubated for 30 minutes at 37°C with RNAse and the DNA was purified by phenol extraction and ethanol precipitation.

To prepare a control template with detectable amounts of randomly ligated DNA fragments, we had to enrich for ligation products of interest (see also results). PCR fragments spanning the restriction sites of interest were gel purified and the DNA concentration was carefully determined using a Cary 100 Bio spectrophotometer (Varian). Equimolar amounts of the different PCR fragments were mixed and digested with the appropriate restriction enzyme followed by ligation. The mix was purified by phenol extraction and ethanol precipitation. The ligated fragments were diluted to the appropriate concentration (see below) and mixed with 300 ng digested and ligated genomic DNA.

PCR analysis of the ligation products.

The linear range of amplification was determined for the fetal liver samples and fetal brain samples by serial dilution. An appropriate amount of DNA within the linear range (typically 300ng of DNA, for both liver and brain) was subsequently used for the experiments. The linear range of the control template was determined with a serial dilution of the random ligation mix made in the same amount (300ng) of digested and ligated genomic DNA. Standardly, the 5' side of each restriction fragment was used to design primers unless this coincided with repetitive DNA sequences. Primer sequences are available on request. PCR products were run on 2% agarose gels and quantified on a Typhoon 9200 imager (Molecular Dynamics). All data points were generated from an average of five (with a minimum of three) different experiments performed in duplo. PCR products of the ligated fragments were run on agarose gels and quantitated. Cross-linking frequencies were calculated using the equation shown in figure 3.1C. All probes (I-XIII) were tested against all other probes. A selection of the results is presented, and data not shown are in agreement.

As shown before [229] formation of ligation products was strictly dependent on both ligation and cross-linking, i.e. lowering the amount of formaldehyde resulted in loss of PCR product, as did the omission of T4 ligase (data not shown).

Acknowledgements

We would like to thank John Strouboulis, Niels Galjart and Sjaak Philipsen for comments on the manuscript. Also, we thank our collegues in the lab for help with the logistics. This work is supported by NWO (the Netherlands Organisation for Scientific Research) to WdL as part of the Innovational Research Incentives Scheme and by NWO and EC grants to FG.

Chapter 4

Structural dynamics of a nuclear compartment of *cis*-regulatory DNA elements and β-globin genes during erythroid differentiation and development.

Structural dynamics of a nuclear compartment of *cis*-regulatory DNA elements and β-globin genes during erythroid differentiation and development.

Robert-Jan Palstra[1], Bas Tolhuis[1], Erik Splinter, Rian Nijmeijer, Frank Grosveld and Wouter de Laat

Department of Cell Biology and Genetics, ErasmusMC, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Fax: +31-10-4089468

[1] These authors contributed equally

Corresponding author: w.delaat@erasmusmc.nl

This chapter has been submitted for publication

Summary

Efficient transcription of genes requires a high local concentration of the relevant *trans*-acting factors. Nuclear compartmentalization can provide an effective means to locally increase the concentration of rapidly moving *trans*-acting factors, and may be achieved by spatial clustering of chromatin-associated binding sites for such factors [75, 108, 152, 272, 273]. Here we analyse the structure of an erythroid-specific spatial cluster of *cis*-regulatory elements and active β-globin genes, the Active Chromatin Hub (ACH) [190], at different stages of development and in erythroid progenitors. We show, in mouse and man, that a core ACH is developmentally conserved and consists of the hypersensitive sites (HS1-6) of the locus control region (LCR), the upstream 5'HS-60/-62 and downstream 3'HS1. Globin genes switch their interaction with this cluster during development, correlating with the switch in their transcriptional activity [19]. In murine erythroid progenitors that are committed to, but do not yet express globin, only the interactions between 5'HS-60/-62, 3'HS1 and HS at the 5' side of the LCR are stably present. Upon induction of differentiation, these sites cluster with the rest of the LCR and the gene that gets activated. We conclude that during erythroid differentiation, *cis*-regulatory DNA elements create a developmentally conserved nuclear compartment dedicated to RNA polymerase II-transcription of β-globin genes.

Results/discussion

The mouse and human β-globin locus contain an upstream LCR and multiple β-like genes arranged from 5' to 3' in order of their developmental expression (figure 4.1A and 4.3A). In addition there are several distal hypersensitive sites (HS), including a downstream 3'HS1 (approx. 20kb 3' of the β genes) and two upstream HS, ~60 kb (mouse) and ~110 kb (human) away from the genes [164]. The loci are embedded in an olfactory receptor gene cluster that is inactive in erythroid cells [162]. To investigate the spatial organisation of β-globin gene loci in mouse and man during development and erythroid differentiation, we applied chromosome conformation capture (3C) technology (see Methods [190, 229]). 3C-technology involves quantitative PCR-analysis of cross-linking frequencies between two given DNA restriction fragments, which gives a measure of their proximity in the nuclear space. Local chromatin configuration has no effect on digestion efficiency, implying that the assay is not biased due to preferential restriction enzyme digestion of one site over the other (figure 4.1D-G; for other controls see Methods and [190]).

First, we determined the spatial organisation of the murine β-globin locus in primitive erythroid cells present in 10.5 dpc embryonic blood, that predominantly express the embryonic εy and βh1 globin genes [217]. Cross-linking frequencies were determined for 66 pairs of *HindIII* restriction fragments, spread over ~170 kb of DNA encompassing the murine β-globin gene cluster. The 3C-measurements indicate a basic structural organisation in primitive cells very similar to that observed previously in definitive blood cells isolated from 14.5 dpc fetal liver [190]. This is best illustrated by comparing the locus-wide cross-linking frequencies of a restriction fragment that contains HS4-5 of the LCR. Two peaks of high cross-linking frequency with this genomic site stand out in primitive blood cells: one with the upstream HS-60/-62 and another with 3'HS1 downstream of the genes (figure 4.1B). Significantly lower cross-linking frequencies were found with fragments in between, suggesting that the LCR interacts with these distal HS through looping. The same interactions were observed in definitive blood cells that exclusively express the adult βmajor and βminor globin genes [217], where βmajor is also found in close proximity (figure 4.1B, and see below). In contrast, in non-expressing brain cells HS4-5 shows no peaks of interaction with distal DNA fragments, suggesting a linear conformation of the transcriptionally inactive locus [190]. Similar

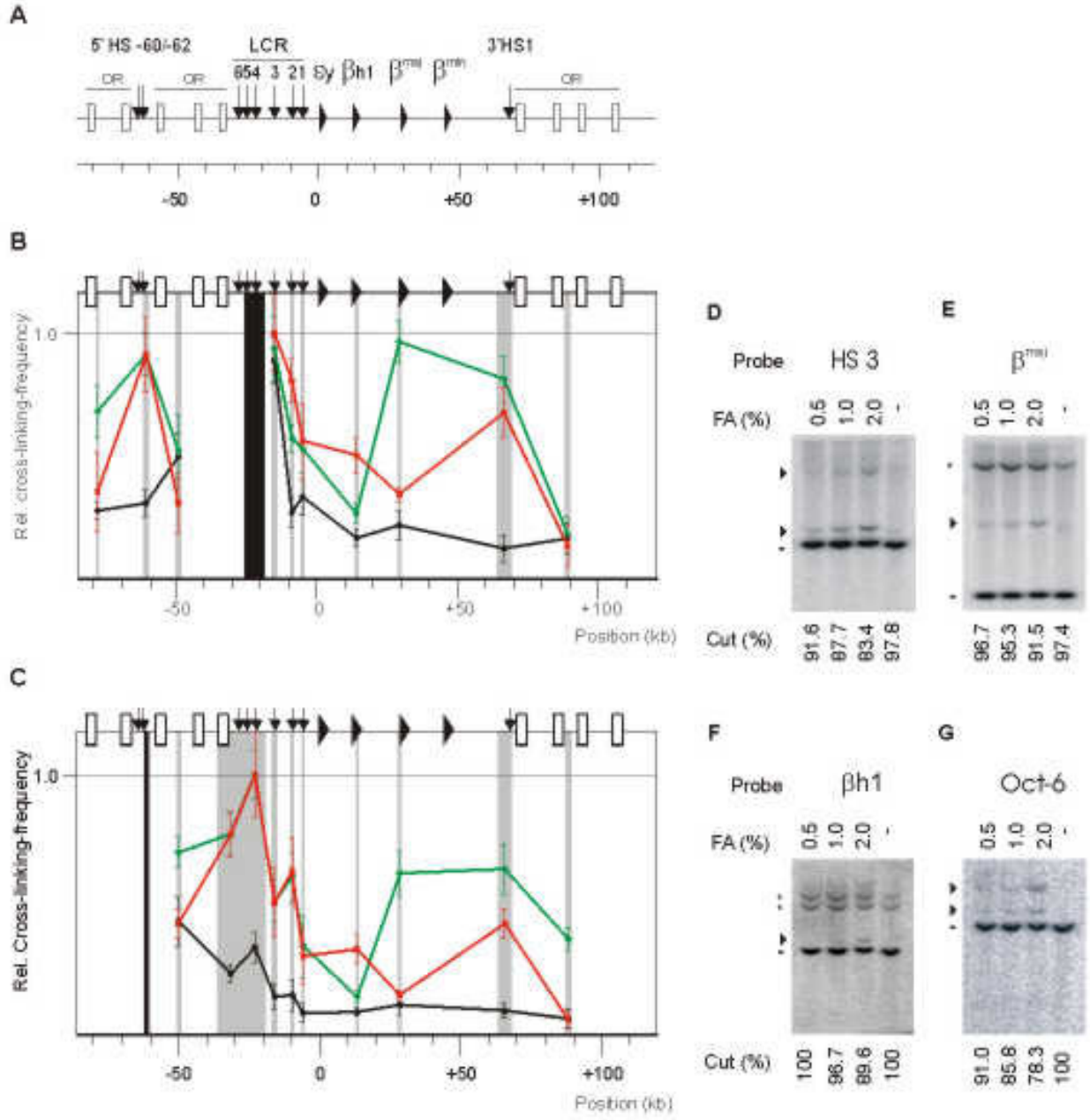


Figure 4.1. Spatial organization of the murine β-globin locus.

Schematic presentation of the mouse locus (A). Arrows depict the individual HSs, globin genes are indicated by triangles, and boxes indicate the olfactory receptor (OR) genes. Erythroid-specific and developmentally stable clustering of *cis*-regulatory elements (B-C). Relative cross-linking frequencies observed in primitive erythrocytes are shown in red, definitive erythrocytes in green, and non-expressing brain in black. Grey shading indicates position and size of the analysed fragments, while black shading represents the 'fixed' fragments HS 4-5 (B) and 5'HS –60/-62 (C). Within each graph, the highest cross-linking frequency value was set to one. The x-axis shows position in the locus.

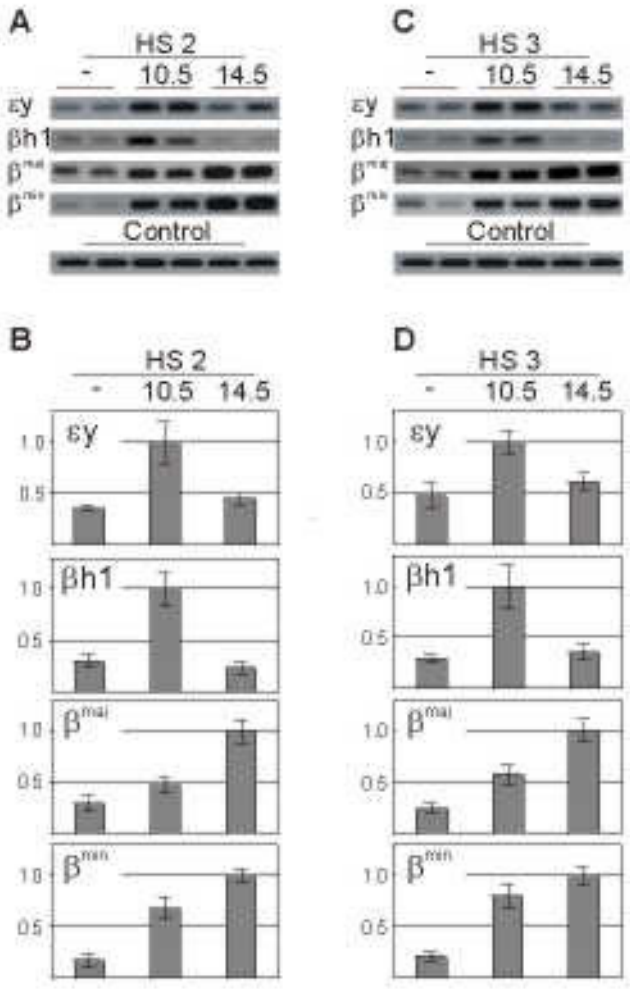


Figure 4.2. A Developmental switch occurs in contacts between individual β-globin genes and the core ACH of the murine β-globin locus.

Cross-linking frequencies of HS2 and the β-globin genes were measured (A-B). An example of PCR amplified ligation products is shown on 2% agarose gel (A), as well as the quantified data of all experiments (at least five, in duplo, per primer set) (B). (C-D) Identical to (A) and (B), but now for HS3 and the β-globin genes. Standard-error-of-mean is indicated. Non-expressing brain is depicted by '-', primitive erythrocytes by '10.5', and definitive erythrocytes by '14.5'. Control is PCR-amplified ligation product of two restriction fragments in the XPB locus (see Methods). Cross-linking frequencies shown in *b* and *d* are not corrected for PCR amplification efficiency and therefore only signals obtained with the same primer set can be compared.

results were obtained when analysing the locus-wide cross-linking frequencies of fragments carrying 5'HS-60/-62 (figure 4.1C) and other HS (data not shown): interactions among the *cis*-regulatory elements of the β-globin locus were found to be conserved between primitive and definitive erythroid cells. We conclude that the *cis*-regulatory elements of the murine β-globin locus spatially cluster to form a transcription regulatory compartment that is conserved between primitive and definitive erythroid cells, two developmentally different types of cells that express a different subset of β-like globin genes. This core ACH includes the two HS at –60 kb, all HS of the LCR and 3'HS1.

(Figure 4.1 continued). Standard-error-of-mean is indicated. Southern blots show that, in definitive erythrocytes digestion efficiency of cross-linked chromatin depends on formaldehyde concentration and is comparable between a hypersensitive site in the LCR (D), a transcribed gene within the locus (E), a non-expressed gene within the locus (F) and a non-expressed gene on a different chromosome (chr. 4) (G). Percentage formaldehyde cross-linking is shown at the top of each blot (- depicts genomic DNA not treated with formaldehyde), while the yield of specifically cut fragments is shown (percentages) at the bottom. Arrowheads depict partial digests and asterisks cross-hybridisation signals with other genes (see Methods).

The main differences in conformation between the two expressing cell types appear to be confined to interactions between the globin genes and the regulatory DNA elements. This is confirmed by measuring cross-linking frequencies with HS2 and HS3 of the LCR, two sites previously shown to be the most prominent transcriptional activating elements [15, 177, 180, 186, 187, 218]. The embryonic globin genes εy and βh1 were found to interact frequently with these elements in primitive erythroid cells, whereas in definitive red blood cells interaction frequencies between these sites dropped to levels similar to what was observed in the inactive brain (figure 4.2A-D). The opposite was seen for the adult βmajor and βminor genes, which interacted most frequently with HS2 and HS3 in definitive erythroid cells. Cross-linking frequencies between these sites in 10.5 dpc embryonic blood were not as low as in brain, probably due the fact that βmajor and βminor are already transcriptionally active at this stage, albeit at less than 10% of the levels observed in definitive cells. Alternatively, it may merely be the result of 3'HS1 interacting with the LCR and the adult genes being dragged along, as we previously found that the region between βminor and 3'HS1, which is full of repetitive sequences, acts as a rigid region [190]. These data demonstrate that there is a developmental switch in contacts between the different globin genes and a core ACH created by regulatory elements that surround the genes *in cis*. This structural change correlates with the developmental switch in expression of the genes.

To further investigate the significance of our findings, we analysed the conformation of the human β-globin locus at different stages of development. The mouse and human β-globin gene loci show a high degree of nucleotide sequence conservation, particularly at regions implicated in gene regulation [162, 274]. We made use of transgenic mice carrying a single copy of a 185 kb PAC (figure 4.3A) spanning the human β-globin locus that displayed a normal expression pattern [199, 232] (Patrinos, in prep.). Although large, this PAC does not include the human equivalent of the murine 5'HS-60/-62, which is located ~110 kb upstream of the human globin genes [162, 164]. We analysed the conformation of the transgenic human globin locus in 10.5 dpc embryonic blood, 14.5 dpc fetal liver and 14.5 dpc fetal brain, measuring almost all of the 120 site pairs that can be formed between the 16 *EcoRI* fragments that were selected for analysis. The locus-wide cross-linking frequencies of a fragment corresponding to 3'HS1 illustrate that also the transgenic human locus forms a core ACH, consisting of the 3'HS1 and the HS of the LCR, that is conserved in primitive and definitive erythroid cells (figure 4.3B). The structural changes we observed primarily concerned the position of the genes relative to this core ACH, correlating with transcriptional activity. Thus, the embryonic ε and the two γ genes most frequently interact with HS2-4 (figure 4.3C-E) and 3'HS1 (figure 4.3B) in primitive erythroid cells and the adult β gene primarily contacts the ACH in definitive cells (figure 4.3C-E). Identical results were found for a *HindIII* digest and for a different transgenic PAC line (data not shown). Results obtained with definitive erythroid cells isolated from adult bone marrow ($Ter119^+$) were identical to those found for 14.5 dpc fetal liver cells (data not shown). It is interesting to also note the decreased cross-linking frequency of HS5 in the definitive cells as we have recently shown that this element has LCR blocking activity in primitive but not definitive erythroid cells [35]. We conclude that the overall spatial organisation of the β-globin gene cluster is conserved from mouse to man.

Next, we determined β-globin genomic site interactions in I/11 erythroid progenitor cells that are committed to, but do not yet express the β-globin genes. If exposed to physiologically relevant stimuli, I/11 cells synchronously undergo the normal *in vivo* differentiation program to mature terminally into enucleated erythrocytes [279, 280]. As expected, in differentiating I/11 cells that actively transcribe the adult β-like globin genes the locus adopts a spatial organisation very similar to what we observed previously in definitive erythroid cells isolated from fetal livers [190] (figure 4.4). However, in uninduced

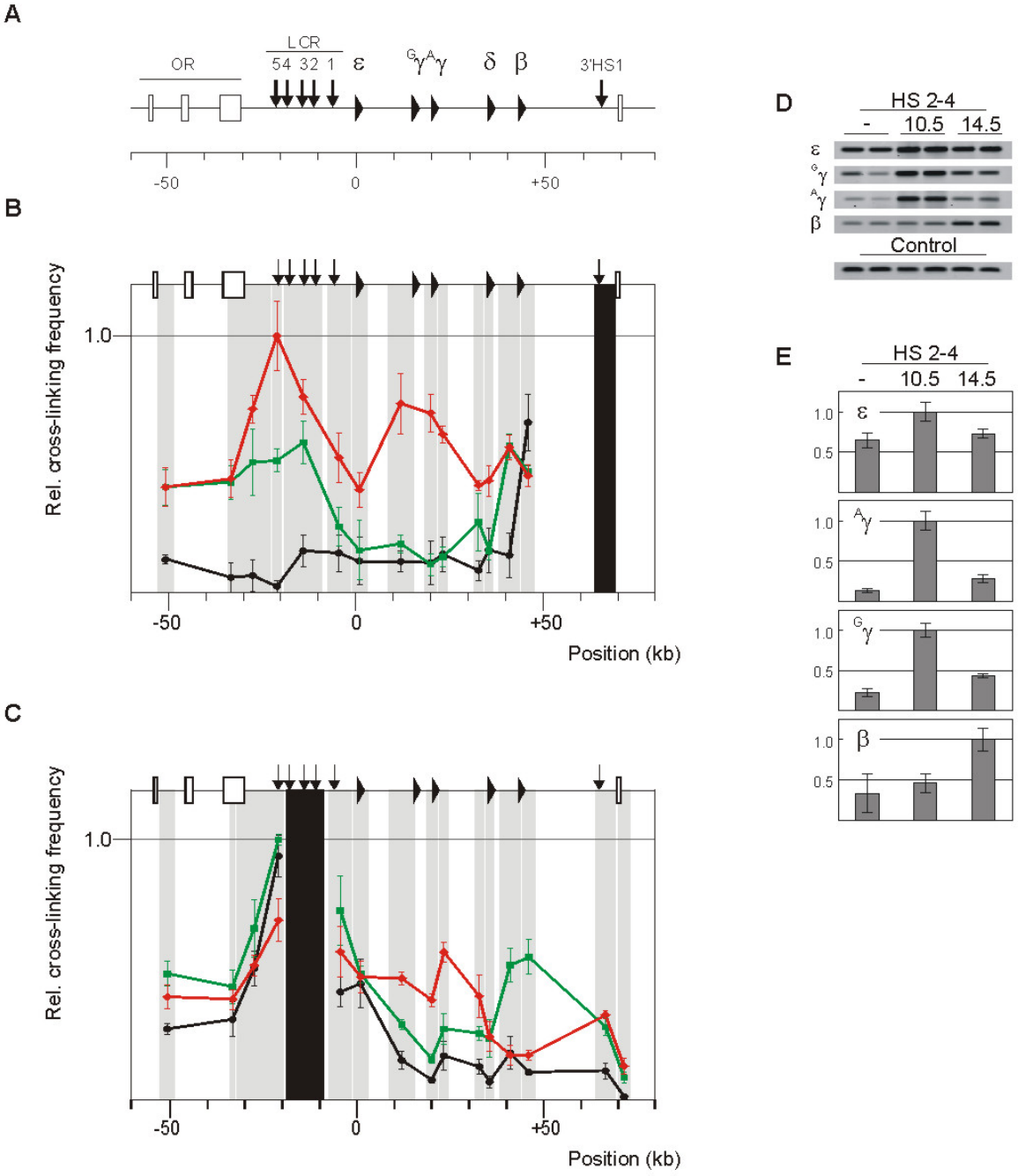


Figure 4.3. Spatial organization of the human β-globin locus.

Controls, symbols, color patterns and numbering are as in figure 1 and 2. Schematic presentation of the human locus (A). Locus-wide cross-linking frequencies of a 3'HS1 fragment show erythroid cell specific clustering with the LCR throughout development (B). Developmental switching in contacts of the LCR between the different β-globin genes as shown by locus-wide cross-linking frequencies of HS2-4 (C). The contacts between the HS2-4 of the LCR and individual β-globin genes alter during development in erythroid cells, as shown (D) by an example on agarose gel and (E) quantified data (at least five experiments in duplo per primer set).

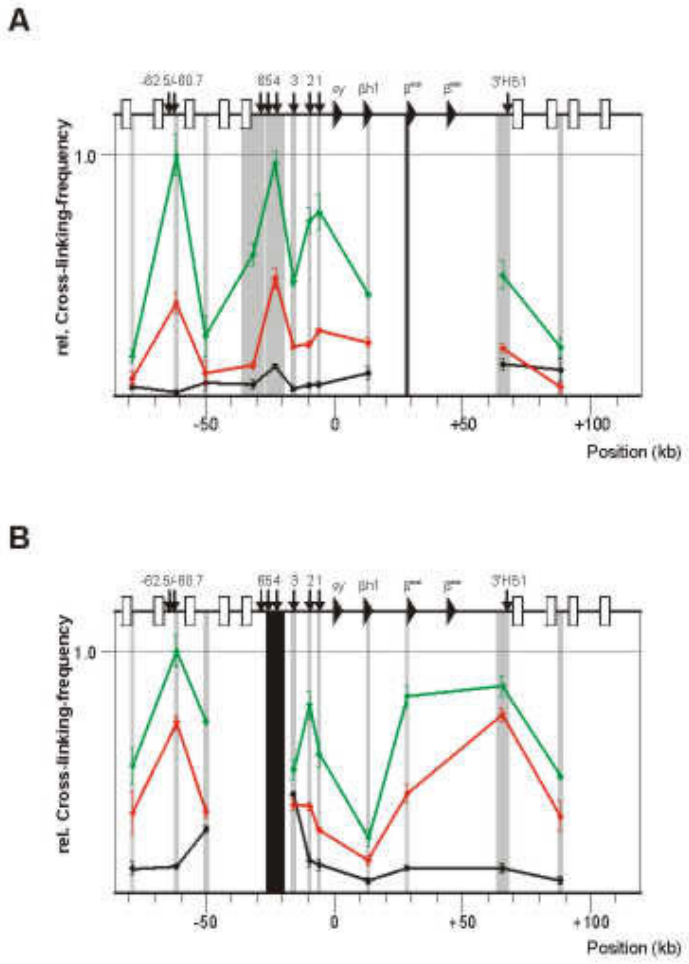


Figure 4.4. Spatial organization of the murine β-globin locus in erythroid progenitors.

Controls, symbols and numbering are as in figure 1, black lines represent brain, red lines proliferating l/11 erythroid progenitor cells and green differentiated l/11 cells. Locus-wide cross-linking frequencies of βmajor (A). Locus wide cross-linking frequencies of HS4-5 (B). Note that only the interactions among HS4-5, HS-60/-62 and 3'HS1 are already fully established in non-expressing progenitor cells.

proliferating l/11 cells that do not yet express the β-globin genes, a different structure is observed. Locus-wide cross-linking frequencies of a fragment corresponding to the βmajor gene were found to be reduced compared to those observed in erythroid cells expressing the gene (figure 4.4A). However, the structure of the locus poised for transcription is clearly different from that of the inactive locus in brain cells. This structure is better resolved by looking at the locus-wide cross-linking frequencies of the restriction fragment that contains HS 4-5 of the LCR. Two peaks of high cross-linking frequency with this fragment stand out in erythroid progenitor cells: one with 5'HS-60/-62 and another with 3'HS1 (figure 4.4B). Interactions among these three sites occur almost as frequently in proliferating progenitors as in differentiating erythroid cells, whereas all other interactions examined between globin site pairs are strongly reduced in progenitor cells (figure 4.4A-B, and data not shown). We conclude that the β-globin locus that is poised for transcription in progenitor cells adopts a looped conformation through frequent interactions between the two distal regulatory elements at either end of the locus (HS-60/-62 and 3'HS1) and HS at the 5' side of the LCR (HS4,5 or 6, we currently cannot say which of these HS is responsible for direct interaction). Upon induction of differentiation, clustering with the active genes and the complete LCR is established and the β-globin genes are being expressed (figure 4.5).

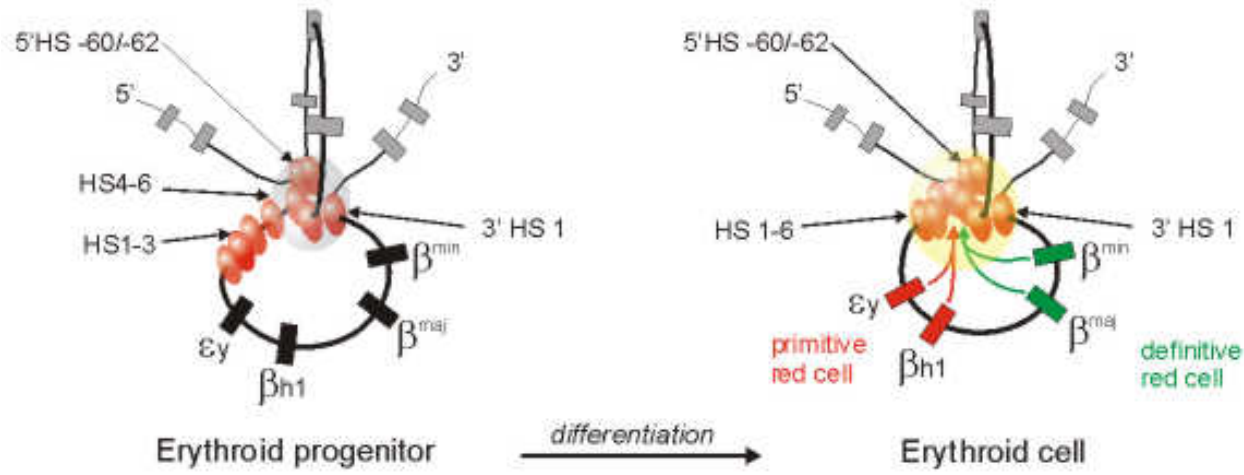


Figure 4.5. *Cis*-regulatory elements of the β-globin locus create a nuclear compartment dedicated to RNA polymerase II transcription: the Active Chromatin Hub.

2D-presentation of 3-dimensional interactions that occur between regulatory DNA elements 130 kb apart (red ovals) and β-globin genes (active: red and green rectangles; inactive: black) in erythroid progenitors (left) and differentiated primitive and definitive erythroid cells (right). In erythroid progenitors not expressing globin, a substructure (grey sphere) is present which is formed through interactions between the upstream 5'HS-60/-62, the downstream 3'HS1 and HS at the 5' side of the LCR (HS4-6; we currently cannot say which of these HS is directly responsible for this interaction). During erythroid differentiation, the β-globin gene that gets activated and the rest of the LCR stably interact with this sub-structure to form a functional ACH (yellow sphere); β-globin gene expression is activated. Clustering of binding sites for transcription factors in the ACH causes local accumulation of cognate proteins and associated positive chromatin modifiers, required to drive efficient transcription of the globin genes. The core of the ACH is erythroid-specific and developmentally stable; a developmental switch occurs in globin genes entering this nuclear compartment, as depicted by the arrows. Inactive globin and olfactory receptor genes (grey squares) loop out.

In summary, our data strongly suggest that regulatory elements surrounding the β-globin genes in *cis* create an erythroid-specific developmentally stable nuclear compartment dedicated to RNA polymerase II transcription (figure 4.5). A sub-structure is already present in erythroid progenitors that do not express globin, and it is worth noting that the three sites involved in this structure all bind CTCF [39]. It should be noted that 3'HS1 and 5'HS-60/-62 are dispensable for normal globin gene expression in transgenic mice [35, 199], suggesting these sites have a more general structural role not related to transcription per se.

Spatial clustering of transcription regulatory elements results in a high local concentration of DNA binding sites for cognate transcription factors, which as a consequence accumulate at the site. Efficiency of transcription is proportional to the concentration of transcription factors involved, and in agreement we found that proximity of β-globin genes to the ACH correlated with transcriptional activity. The paradigm of a chromatin-associated nuclear compartment is the nucleolus, dedicated to RNA polymerase I transcription of ribosomal RNA genes [108]. No pol II-dependent gene-specific compartments have been described before, but a precedent for this was provided by electron microscopy studies showing that RNA polymerase II clusters in discrete transcription factories in the nucleus [110, 275]. The fact that the density of RNA polymerases on active β-globin and ribosomal RNA genes is much higher than on most other active genes [111, 112] suggests that such nuclear compartments formed by numerous chromatin-associated regulatory elements primarily function to increase the efficiency of transcription.

Methods

Chromosome Conformation Capture (3C).

Isolation and formaldehyde fixation of primary cells, restriction enzyme digestion of cross-linked DNA in the nucleus, intramolecular ligation, reversal of cross-links, PCR-analysis of ligation products and calculation of relative cross-linking frequencies was done as described before [190, 229], with some modifications. Prior to fixation, cells obtained from embryonic blood (10.5dpc embryos), fetal liver and fetal brain (both 14.5dpc embryos) ($4x10^7$ cells per tissue) were forced through a cell-strainer cap (Falcon #352235) to obtain a homogeneous single cell suspension. To correct for differences in quality and quantity of template, we normalize ligation frequencies between globin site pairs to those detected between two restriction fragments (with the sites analysed 8.3 kb apart) in the XPB locus (instead of the previously used CalR locus [190]). XPB encodes a subunit of the basal transcription factor TFIIH, and we assumed that expression levels and spatial conformation of this gene are similar in all analysed tissues. To be able to compare signal intensities obtained with different primer sets in a quantitive manner, a control template is included containing all possible ligation products in equimolar amounts to correct for the PCR amplification efficiency of each set. For this purpose we used BAC and PAC clones spanning the complete loci (instead of the previously used PCR fragments that span the restriction sites of interest [190]). For the mouse β-globin locus we used a 214 kb BAC (#RP23-370E12, Ensembl Genome Browser, http://www.ensembl.org), and for the human β-globin locus we used a 185 kb PAC [232]. In addition, we used a 60-70 kb PAC containing the mouse XPB locus (PAC Clone #443-C18, MRC gene service, http://www.hgmp.mrc.ac.uk). We mixed either the mouse globin BAC or the human globin PAC with the XPB PAC at equimolar amounts. Subsequently, the mixes were digested and ligated as described [190]. We could not obtain control PCR products with primers designed to analyse fragments containing εy and βminor, due to polymorphisms in the BAC clone #RP23-370E12. As a consequence, these fragments were not included in the locus-wide cross-linking frequency analysis (see figure 4.1).

Southern blotting.

Fetal liver cells (14.5dpc embryos) were treated as above (with indicated formaldehyde concentrations), but ligation was omitted and 10 μg of purified DNA was analysed by southern blotting. The following probes were used: βh1, a 255 bp *Hinf*I fragment, hybridises to a 2.7 kb *Hind*III βh1 fragment and βh0 (5.5 kb) and βh2 (6.4 kb) psuedogene fragments; βmajor, a 700 bp *Hind*III/*Nco*I fragment, hybridises to a 1.0 kb *Hind*III fragment and a βminor (8.6 kb) fragment; HS3, 300 bp PCR fragment, hybridises to a 2.0 kb *Hind*III fragment; Oct-6, 100 bp PCR fragment, hybridises to a 4.0 kb *Hind*III fragment.

Cell culture.

I/11 cells were culture as described previously [279, 280]. Briefly, proliferating I/11 cells were maintained in StemPro-34™ containing 2 units/ml human recombinant erythropoietin, 100 ng/ml murine recombinant SCF, 10^{-6} M dexamethasone and 40 ng/ml insulin-like growth factor. Cells were expanded by daily partial medium changes and addition of fresh factors, keeping cell density between 1.5–4 $x10^6$ cells/ml. For induction of differentiation, continuously proliferating I/11 cells were removed from the culture, washed twice in PBS, and seeded at 2–3 $x10^6$ cells/ml in differentiation medium containing 10 units/ml Epo,
4 $x10^{-4}$ IE/ml Insulin, the Dex-antagonist ZK-112993 (3 $x10^{-6}$ M), and 1 mg/ml iron-saturated human transferrin. Differentiating erythroblasts were maintained at densities between 2-6 $x10^6$ cells/ml. For 3C-analysis of differentiating I/11 cells, cells were fixed with formaldehyde 40 hours after induction and processed as described above.

Acknowledgements

We thank Karim Hussain and Rudi Hendriks for technical assistance and Marieke von Lindern for her kind gift of I/11 cells. This work is supported by NWO (The Netherlands Organisation for Scientific Research) to WdL as Part of the Innovational Research Incentives Scheme and by NWO and EC grants to FG.

Chapter 5

General discussion: the active chromatin hub

The active chromatin hub (ACH) represents a particular spatial organization of the β-globin loci in erythroid cells that includes clustering of DNase I HSs, i.e. *cis*-regulatory DNA elements. Clustering of sites in mouse β-globin locus occurs between the actively transcribed genes and all sites of the LCR together with two distal hypersensitive regions. In contrast, the intervening chromatin that contains inactive genes does not participate but loops out (chapter 3) [190]. The LCR and distal hypersensitive regions form a developmentally stable core ACH. The globin genes switch their interaction with this cluster during development, correlating with the switch in their transcriptional activity (see figure 4.5). A similar spatial conformation was found with a human β-globin locus in transgenic mice. In murine erythroid progenitors that are committed to, but do not yet express globin, only the interactions between 5'HS-60/-62, 3'HS1 and HS at the 5' side of the LCR are stably present (see figure 4.5). Upon induction of differentiation, these sites cluster with the rest of the LCR and the gene that gets activated. In conclusion, we propose that during erythroid differentiation, *cis*-regulatory DNA elements create a developmentally conserved nuclear compartment dedicated to RNA polymerase II-transcription of the β-globin genes (chapter 4) [191]. The spatial clustering of regulatory elements (ACH) results in a high local concentration of DNA binding sites for cognate transcription factors and their interacting partners (e.g. HATs, chromatin remodeling enzymes), which consequently accumulate at the site. Efficiency of transcription is proportional to the concentration of transcription factors involved, and in agreement we found that proximity of β-globin genes to the ACH correlated with transcriptional activity. Below several speculative properties of the ACH have been put forward to explain its role in efficient transcription of the globin genes.

Formation and maintenance

Formation of the ACH may depend on affinity between *cis*-regulatory DNA elements and their cognate binding transcription factors (in cooperation with other *trans*-acting factors). For instance, EKLF binds with higher affinity to the adult β-globin promoter than to the γ-globin promoter [276]. Furthermore, changes in EKLF concentration induced alterations at the chromatin structure of the β-globin promoter and 5' HS3 with coincident absence of β-globin gene transcription. These chromatin alterations at the β-promoter coincide with upregulated γ-globin expression in the fetal liver [181]. This suggests that modulation of the *trans*-acting environment will influence ACH formation and interactions. Indeed, 3C measurements showed that absence of EKLF leads to alterations in the spatial interactions of the mouse β-globin locus (Roy Drissen, personal comm.). In this context, it would be interesting to examine the effects on ACH interactions of other erythroid cell-specific transcription factors. In a Friend-virus erythroleukemia cell line, expression of the p45 subunit of NF-E2 is undetectable due to proviral integration in one allele and loss of the other allele. The complete loss of p45 in this cell line is associated with a drastic reduction in expression of α- and β-globin genes [277, 278]. These cells would be suitable to examine the spatial interactions of the enhancer property of 5' HS2, which depends on a tandem NF-E2 binding site that is not found as a tandem array in any of the other HSs [179]. GATA-1 binding sites are present in globin gene promoters and the HSs of the LCR. This factor interacts with several complexes that either activate or repress transcription (Patrick Rodriguez, personal comm.). Thus, altering GATA-1 concentrations in erythroid cells or expressing mutant GATA-1 proteins might give valuable insights in ACH interactions. In addition to erythroid cell-specific factors, other more widely expressed factors may be participating in the ACH. One candidate would be the vertebrate enhancer-blocking factor CTCF. The human and mouse loci contain several putative CTCF binding sites that are present as homologous motifs in the distal hypersensitive regions and 5' HS5 of the LCR. Indeed, ChIP analyses with a CTCF antibody showed that in erythroid tissue 5' HS-62/-60, 5' HS5, and 3' HS1 of the mouse locus were signigicantly enriched [39, 184]. The 3C measurements demonstrated that spatial interactions between the LCR and the distal hypersensitive

regions occurred most frequently with a genomic site containing both 5' HS5 and HS4 of the mouse β-globin locus (see chapter 3 and 4) [190, 191]. In mouse erythroid progenitor cells [279, 280], which do not transcribe the globin genes yet, interactions between 5' HS5/4 of the LCR and the distal regions is still observed, while the frequency of other interactions are reduced. Upon differentiation, a complete ACH is in mature erythroid cells formed (chapter 4). Likewise, spatial interactions between the human 5' HS5 and 3' HS1 appeared to occur with high frequency. In primitive erythroid cells, these interactions occurred at higher frequencies than those between 5' HS5 and other genomic sites (chapter 4) [191]. Interestingly, the human 5' HS5 was found to have insulating properties in primitive cells, but not definitive cells [35]. In addition, a colony assay with 3' HS1 in beween the 5' HS2 enhancer and a reporter gene resulted in a 3-fold reduction of expression compared to a control construct without the 3' HS1. The authors interpreted this as an enhancer-blocking activity of 3' HS1 [39]. Thus, it would be tempting to speculate that CTCF may play a role in ACH activity. However, the precise nature of its function needs to be investigated. Currently (conditional) knock-out studies of CTCF are in progress that would provide an excellent tool. A second group of ubiquitous factors may participate in the ACH in a slightly different way and were originally identified in *Drosophila*. These factors were shown to facilitate enhancer-promoter communication and proposed to cooperate with different LIM domain proteins and other factors [225, 265]. These so-called facilitators have mammalian homologues LIM-domain binding protein 1 (Ldb1) and Idn3 [265, 266, 281]. The Ldb-1 protein binds nuclear LIM-domain proteins and was shown to form a large complex with erythroid cell-specific transcription factors, including GATA-1 [282, 283]. This complex was tethered to the *c-kit* promoter via a specificity protein 1 (Sp1) motif, through direct interactions between elements of the complex and the Sp1 zinc finger protein [283]. Thus, these facilitators act through interactions with other *trans*-acting factors and may serve as nucleoprotein structures that link the interacting *cis*-regulatory DNA elements together. However, (conditional) knock-out studies of the genes encoding facilitators, which are in progress, should address these hypothetical interactions in mammalian cells.

Although DNase I HS formation precedes transcription, hypersensitivity of the LCR depends on the presence of active promoters [168-171], supporting the idea that stability and maintenance of sites may rely on ACH formation. Deletions in/of LCR elements results in variegated expression. Importantly, the affected loci are DNase I insensitive in the nontranscribing portion of the cells [27, 28]. Moreover, alterations in the *trans*-acting environment can modulate variegated expression patterns [258, 284]. In the light of the ACH, we propose that modifying the number of *cis*-regulatory DNA elements and/or their bound *trans*-acting factors may alter existing interactions and consequently the expression levels of the genes. This would be analogous to the self-organization capacities of nuclear compartments [152]. For instance, introduction of extrachromosomal rRNA genes triggers the spontaneous formation of novel nucleoli [102, 108].

The initial formation of DNase I HSs is largely unknown, but may be explained by a mass-action model [220, 259]. For example, formation of LCR sites depends on multiple transcription factors [173, 174, 285] and position-independent expression of an enhancer driven construct is only observed with multi-copy integrations [15]. This suggests that a critical number of regulatory elements and bound factors are required for HS formation. A mass action model strongly depends on a critical number of interactions to increase the likelihood of establishing a stable structure and shows similarities with the stop-and-go properties of *trans*-acting factors that are involved in nuclear compartmentalization.

Chromatin opening and the ACH

How do *cis*-regulatory elements and *trans*-acting factors collaborate efficiently on the restrictive chromatin template to drive RNAP II transcription? Erythroid cell-specific transcription factors (e.g. EKLF, NF-E2, and GATA-1) not just bind the *cis*-regulatory DNA elements of the locus, but also interact with numerous co-factors, such as HATs and chromatin remodeling enzymes. They may be essential for targeting these other *trans*-acting factors to specific *cis*-regulatory DNA elements. Indeed, *cis*-regulatory DNA elements that participate in the ACH contain hyperacetylated histones [39, 193, 195, 198]. Histone hyperacetylation is a biochemical 'mark' of transcriptionally active euchromatin. In addition, local hyperacetylation may be bound by additional *trans*-acting factors, such as SWI/SNF complexes, which have a bromodomain that can bind to acetylated histones [53, 61]. Thus, these chromatin modifying co-factors may contribute in the establishment and maintenance of the open chromatin state. Immunofluorescence detection of these co-factors in conjunction with detection of ongoing globin gene transcription should determine whether these factors accumulate locally at active globin transcription sites. If so, this would support the concept of the ACH nuclear compartment (figure 5.1). Alternatively, disruption of a continuous array of nucleosomes can cause a barrier for the spreading of silent chromatin (nucleosomal gap model) [286]. Nucleosome disruption at the most distal *cis*-regulatory DNA elements present in the ACH may function in parallel with targeted chromatin modifying co-factors to counteract silencing [20].

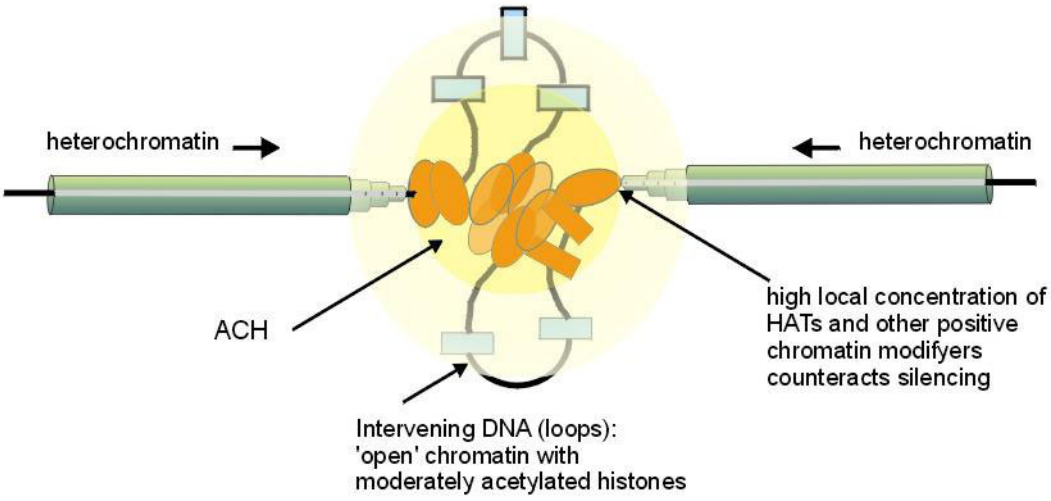


Figure 5.1. Hypothetical: the ACH and transcription in repressive chromatin.

2D presentation of a 3D gene cluster embedded in heterochromatin (green), with active genes and cognate *cis*-regulatory DNA elements in red and inactive genes in blue. The cluster of regulatory sequences and active genes forms an ACH, which harbors a high local concentration (indicated by bright yellow inner circle) of transcription factors, HATs and other positive chromatin modifying factors. Spreading of heterochromatic silencing (green) is blocked at the ACH, mainly due to the local accumulation of these factors. Concentration of positive factors decreases with increasing distance from the ACH (indicated by the outer light-yellow circle). Loops of intervening DNA (with inactive genes) protrude from the ACH; the nearby hub of positive chromatin modifiers affects the chromatin status of these loops. This figure was adapted from [20].

The overall DNase I sensitivity and the moderately increased acetylation levels of the locus were proposed to reflect the existence of chromosomal domains [2, 198, 261]. However, recent data do not appear to conform to existing models of domain formation and demonstrate an unexpected complexity associated with the active β-globin locus [39]. Alternatively, the observed overall chromatin modulations can also be byproducts of activities concentrated at the ACH without necessarily having functional significance [20]. The alternating globin transcription [111] implies that the ACH is a dynamic structure, like many other nuclear compartments. Kinetic studies have demonstrated that proteins enter and exit nuclear compartments at high rates and interact with chromatin in a stop-and-go fashion (see chapter 1, paragraph 1.3 [150]). Like this, co-factors that enter and exit the ACH may shortly and aspecifically interact with chromatin that is present in the looped out DNA. The frequent occurrence of insulator elements at the edges of chromosome domains was interpreted as boundaries of these domains [36]. However, we propose that the ACH relies on dynamic but stable *cis*-regulatory DNA element interactions, rather than the presence of insulating borders, to establish an open chromatin state that allows RNAP II transcription. A concept that is supported by the fact that not all mammalian loci are distinct domains. Instead, several different types of genes may locate immediately adjacent to each other (see figure 1.2.2) [20]. Since the word 'domain' intuitively suggests a physical entity, de Laat and Grosveld [20] proposed to use the term 'functional expression module' to describe a gene or gene cluster and the regulatory elements spread in *cis* that are required for autonomous expression of that particular gene. The frequent occurrence of insulator elements at the edges of gene loci may be explained by evolutionary selection against positioning of insulators within a gene locus [10].

Enhancer-promoter specificity

The observed alternate transcription of the globin genes [111, 216, 217] also implies that only one gene can interact with the core ACH at any given time. Developmental stage-specific negative regulators are thought to silence the human ε- and γ-globin genes at the adult stages [19, 208, 209] and embryonic globin genes are looped out of the ACH in definitive erythroid cells (chapter 3 and 4) [190, 191]. Thus, ε- and γ-promoter-bound silencing complexes may reduce their affinity for the other ACH elements. In addition, the β-globin LCR does not influence transcription of the adjacent OR genes [162, 226] and the OR genes are also looped out of the ACH (see chapter 3 and 4) [190, 191]. Hence, we think that spatial interactions between the ACH and an active gene may depend on affinity between the core ACH bound factors and the factors bound to the gene-proximal *cis*-regulatory DNA elements. This will result in so-called enhancer-promoter specificity.

The relative distance between enhancer and promoter elements, based on gene order, is important for expression levels of the genes [213-215]. The 3C measurements showed that in primitive erythroid cells the proximal embryonic genes enter the core ACH more frequently than the distal adult genes (chapter 4) [191]. Even though, the promoters of all genes have potentially active chromatin structures [8, 195]. Thus, relative distance may determine competition between globin genes for entering the ACH, hence providing another mechanism for enhancer-promoter specificity.

Many collections of *cis*-regulatory DNA elements have been found that meet the functional definition of an LCR (reviewed in [18]). In addition, many genes have to be regulated in a tissue-specific and developmentally manner. Furthermore, some gene loci were found to contain multiple genes that not all relied on the same *cis*-regulatory DNA elements for their transcription (see figure 1.2.2; reviewed in [20]). For instance, the human α-globin locus lies immediately adjacent to several ubiquitously expressed genes. Its major regulatory element resides in an intron of one of these housekeeping genes [287]. Nevertheless, the α-globin genes are tissue-specifically expressed. Therefore, it is tempting to speculate that the formation of an ACH is not just confined to the β-globin

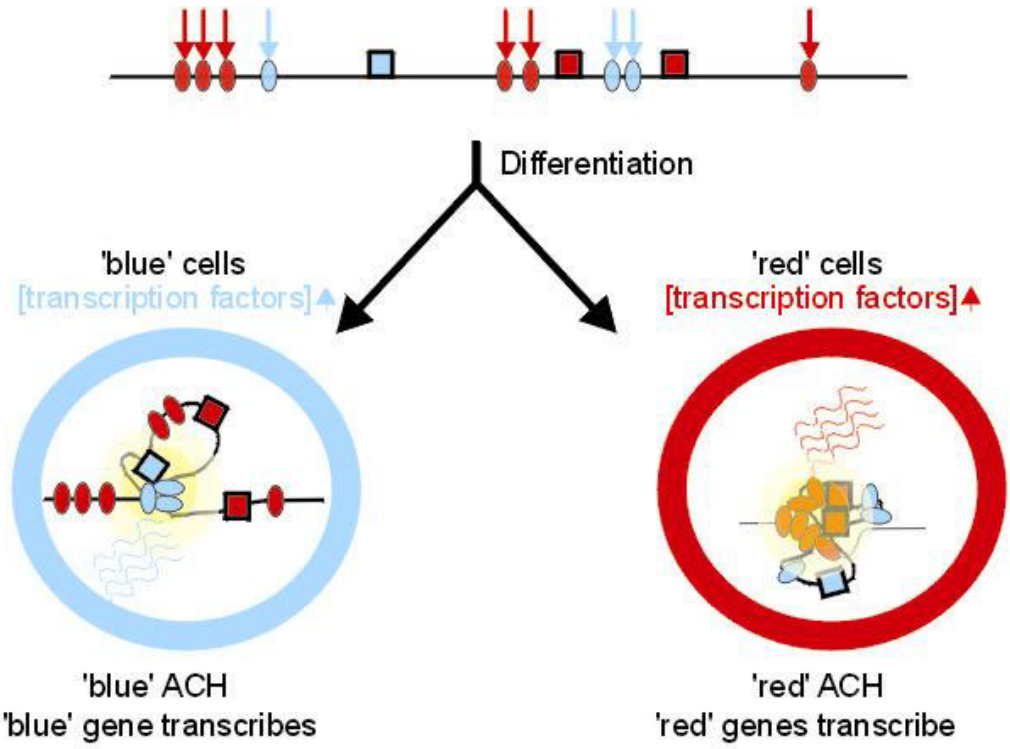


Figure 5.2. Hypothetical: the ACH and expression of overlapping gene loci.

Two virtual gene loci (red and blue) are presented that are differentially regulated and show overlap, with *cis*-regulatory DNA elements as ellipses and genes as boxes. Differentiation to cells with sufficient levels of transcription factors binding to cognate 'blue regulatory elements' ('blue transcription factors' high) will allow the formation of a 'blue ACH', resulting in expression of the 'blue gene'. Similar mechanism applies to the 'red cells' and 'red genes'. This figure was adapted from [20].

loci but may occur at many gene loci in the nucleus (figure 5.2) in order to drive efficient and tissue-specific RNAP II transcription. However, we should keep in mind that transcription of globin genes, as well as a few other genes, is carried out at exceptionally high rates compared to the average gene [111, 112]. The rRNA genes are another set of genes that are highly transcribed, which depends on a nuclear compartment, the nucleolus. Thus, the β-globin and rRNA genes may not be representative of transcriptional regulation of a typical mammalian gene. Instead, the high density of RNAP II complexes on active β-globin and rRNA genes may also suggest that nuclear compartments primarily function to increase the efficiency of transcription.

The 3C analyses represent steady-state average levels, while most interactions in the nucleus are highly dynamic. The ongoing development of a dual tagging system, consisting of Lac^{O}/LacR-CFP and Tet^{O}/TetR-YFP (see chapter 2), would eventually provide an excellent tool to follow the dynamic interactions of the locus in living cells. Of course, changes in *trans*-acting environment could be used to determine the stability of interactions. In addition, this system could be used to analyze the nuclear localization relative to other nuclear compartments, such as transcription factors, chromatin modifying co-factors, and active transcription sites [105, 110].

In potential, 3C can measure cross-linking frequencies throughout a gene locus with the position of restriction sites as the only limitation. Now, the experimental conditions allow analysis with only a few restriction enzymes, all with a 6 bp consensus sequence. Nonetheless, the restriction enzyme is a critical determinant in the 3C analysis. For instance, it is not always possible to obtain restriction fragments containing a single *cis*-regulatory DNA element. When multiple elements are located on one restriction fragment one can no longer determine the individual interactions of these elements. In addition, the size of restriction fragments influences background cross-linking, i.e. larger fragments display higher backgrounds. Ideally, one would like to analyze small restriction fragments that are equal in size, e.g. with a 4 bp consensus sequence. Thus, the 3C conditions are currently adjusted to favor digestion with other restriction enzymes.

It has been proposed that RNAP II genes may be clustered in the nucleus in so-called 'transcription factories'. This was based on the observation that the nucleus produces many more nascent transcripts than there are active transcription sites visible at any given time, while many genes contain only engaged transcription complex [105, 110, 112]. In addition, an erythroid nucleus forms large blocks of heterochromatin upon differentiation. Therefore, these cells may form macromolecular structures containing a cluster of actively transcribed genes, which would be essential for cell survival. One way to address this hypothesis is to clone cross-linked chromatin fragments by using β-globin fragments as bait. In addition, one could examine differences in clustered genes by comparing expressing tissues with nonexpressing tissues. Furthermore, one could follow gene cluster formation during erythroid cell differentiation.

References

1. Struhl K: Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 1999, 98:1-4.
2. Weintraub H, Groudine M: Chromosomal subunits in active genes have an altered conformation. *Science* 1976, 193:848-856.
3. Reeves R: Transcriptionally active chromatin. *Biochim Biophys Acta* 1984, 782:343-393.
4. Blackwood EM, Kadonaga JT: Going the distance: a current view of enhancer action. *Science* 1998, 281:61-63.
5. Albright SR, Tjian R: TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* 2000, 242:1-13.
6. Featherstone M: Coactivators in transcription initiation: here are your orders. *Curr Opin Genet Dev* 2002, 12:149-155.
7. Kollias G, Wrighton N, Hurst J, Grosveld F: Regulated expression of human A gamma-, beta-, and hybrid gamma beta-globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell* 1986, 46:89-94.
8. Groudine M, Eisenman R, Gelinas R, Weintraub H: Developmental aspects of chromatin structure and gene expression. *Prog Clin Biol Res* 1983, 134:159-182.
9. de Villiers J, Olson L, Banerji J, Schaffner W: Analysis of the transcriptional enhancer effect. *Cold Spring Harb Symp Quant Biol* 1983, 47:911-919.
10. Dillon N, Sabbattini P: Functional gene expression domains: defining the functional unit of eukaryotic gene regulation. *Bioessays* 2000, 22:657-665.
11. Martin DI: Transcriptional enhancers--on/off gene regulation as an adaptation to silencing in higher eukaryotic nuclei. *Trends Genet* 2001, 17:444-448.
12. Walters MC, Fiering S, Eidemiller J, Magis W, Groudine M, Martin DI: Enhancers increase the probability but not the level of gene expression. *Proc Natl Acad Sci U S A* 1995, 92:7125-7129.
13. Kollias G, Hurst J, deBoer E, Grosveld F: The human beta-globin gene contains a downstream developmental specific enhancer. *Nucleic Acids Res* 1987, 15:5739-5747.
14. Behringer RR, Hammer RE, Brinster RL, Palmiter RD, Townes TM: Two 3' sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. *Proc Natl Acad Sci U S A* 1987, 84:7056-7060.
15. Ellis J, Talbot D, Dillon N, Grosveld F: Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. *Embo J* 1993, 12:127-134.
16. Grosveld F, Antoniou M, van Assendelft GB, de Boer E, Hurst J, Kollias G, MacFarlane F, Wrighton N: The regulation of expression of human beta-globin genes. *Prog Clin Biol Res* 1987, 251:133-144.
17. Bonifer C: Developmental regulation of eukaryotic gene loci: which cis-regulatory information is required? *Trends Genet* 2000, 16:310-315.
18. Li Q, Peterson KR, Fang X, Stamatoyannopoulos G: Locus control regions. *Blood* 2002, 100:3077-3086.
19. Stamatoyannopoulos G, Grosveld F: Hemoglobin switching. In: *The molecular basis of blood diseases* Edited by Stamatoyannopoulos G, Majerus P, Perlmutter R, Varmus H. pp. 135-182. Philadelphia, PA: W.B. Saunders; 2001: 135-182.
20. de Laat W, Grosveld F: Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* 2003, in press.
21. Magram J, Chada K, Costantini F: Developmental regulation of a cloned adult beta-globin gene in transgenic mice. *Nature* 1985, 315:338-340.
22. Kioussis D, Vanin E, deLange T, Flavell RA, Grosveld FG: Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* 1983, 306:662-666.
23. Driscoll MC, Dobkin CS, Alter BP: Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proc Natl Acad Sci U S A* 1989, 86:7470-7474.
24. Epner E, Reik A, Cimbora D, Telling A, Bender MA, Fiering S, Enver T, Martin DI, Kennedy M, Keller G, Groudine M: 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* 1998, 2:447-455.
25. Reik A, Telling A, Zitnik G, Cimbora D, Epner E, Groudine M: 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* 1998, 18:5992-6000.
26. Bender MA, Bulger M, Close J, Groudine M: Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol Cell* 2000, 5:387-393.
27. Festenstein R, Tolaini M, Corbella P, Mamalaki C, Parrington J, Fox M, Miliou A, Jones M, Kioussis D: Locus control region function and heterochromatin-induced position effect variegation. *Science* 1996, 271:1123-1125.

28. Milot E, Strouboulis J, Trimborn T, Wijgerde M, de Boer E, Langeveld A, Tan-Un K, Vergeer W, Yannoutsos N, Grosveld F, Fraser P: Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* 1996, 87:105-114.
29. Kellum R, Schedl P: A position-effect assay for boundaries of higher order chromosomal domains. *Cell* 1991, 64:941-950.
30. Kellum R, Schedl P: A group of scs elements function as domain boundaries in an enhancer- blocking assay. *Mol Cell Biol* 1992, 12:2424-2431.
31. West AG, Gaszner M, Felsenfeld G: Insulators: many functions, many mechanisms. *Genes Dev* 2002, 16:271-288.
32. Vazquez J, Schedl P: Deletion of an insulator element by the mutation facet-strawberry in Drosophila melanogaster. *Genetics* 2000, 155:1297-1311.
33. Chung JH, Whiteley M, Felsenfeld G: 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* 1993, 74:505-514.
34. Bell AC, West AG, Felsenfeld G: The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 1999, 98:387-396.
35. Wai AW, Gillemans N, Raguz-Bolognexi S, Pruzina S, Zafarana G, Meijer D, Philipsen S, Grosveld F: Characterization of HS5 of the human beta-globin Locus Control Region: a developmental stage-specific border in erythroid cells. *EMBO J* 2003, in press.
36. Bell AC, Felsenfeld G: Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev* 1999, 9:191-198.
37. Chung JH, Bell AC, Felsenfeld G: Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A* 1997, 94:575-580.
38. Pikaart MJ, Recillas-Targa F, Felsenfeld G: Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* 1998, 12:2852-2862.
39. Bulger M, Schubeler D, Bender MA, Hamilton J, Farrell CM, Hardison RC, Groudine M: A Complex Chromatin Landscape Revealed by Patterns of Nuclease Sensitivity and Histone Modification within the Mouse beta-Globin Locus. *Mol Cell Biol* 2003, 23:5234-5244.
40. Woychik NA, Hampsey M: The RNA polymerase II machinery: structure illuminates function. *Cell* 2002, 108:453-463.
41. Hartzog GA: Transcription elongation by RNA polymerase II. *Curr Opin Genet Dev* 2003, 13:119-126.
42. Gregory PD, Schmid A, Zavari M, Munsterkotter M, Horz W: Chromatin remodelling at the PHO8 promoter requires SWI-SNF and SAGA at a step subsequent to activator binding. *Embo J* 1999, 18:6407-6414.
43. Reinke H, Gregory PD, Horz W: A transient histone hyperacetylation signal marks nucleosomes for remodeling at the PHO8 promoter in vivo. *Mol Cell* 2001, 7:529-538.
44. Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M: Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* 2000, 103:843-852.
45. Agalioti T, Lomvardas S, Parekh B, Yie J, Maniatis T, Thanos D: Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. *Cell* 2000, 103:667-678.
46. Lomvardas S, Thanos D: Nucleosome sliding via TBP DNA binding in vivo. *Cell* 2001, 106:685-696.
47. Soutoglou E, Viollet B, Vaxillaire M, Yaniv M, Pontoglio M, Talianidis I: Transcription factor-dependent regulation of CBP and P/CAF histone acetyltransferase activity. *Embo J* 2001, 20:1984-1992.
48. Hatzis P, Talianidis I: Dynamics of enhancer-promoter communication during differentiation- induced gene activation. *Mol Cell* 2002, 10:1467-1477.
49. Strahl BD, Allis CD: The language of covalent histone modifications. *Nature* 2000, 403:41-45.
50. Sterner DE, Berger SL: Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev* 2000, 64:435-459.
51. Peterson CL: HDAC's at work: everyone doing their part. *Mol Cell* 2002, 9:921-922.
52. Kouzarides T: Histone methylation in transcriptional control. *Curr Opin Genet Dev* 2002, 12:198-209.
53. Jenuwein T, Allis CD: Translating the histone code. *Science* 2001, 293:1074-1080.
54. Boyes J, Byfield P, Nakatani Y, Ogryzko V: Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* 1998, 396:594-598.
55. Zhang W, Bieker JJ: Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases. *Proc Natl Acad Sci U S A* 1998, 95:9855-9860.
56. Chen CJ, Deng Z, Kim AJ, Blobel GA, Lieberman PM: Stimulation of CREB binding protein nucleosomal histone acetyltransferase activity by a class of transcriptional activators. *Mol Cell Biol* 2001, 21:476-487.
57. Peterson CL: Chromatin remodeling: nucleosomes bulging at the seams. *Curr Biol* 2002, 12:R245-247.
58. Liu R, Liu H, Chen X, Kirby M, Brown PO, K. Z: Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* 2001, 106:309-318.
59. Krebs JE, Fry CJ, Samuels ML, Peterson CL: Global role for chromatin remodeling enzymes in mitotic gene expression. *Cell* 2000, 102:587-598.

60. Ahringer J: NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet* 2000, 16:351-356.
61. Kingston RE, Narlikar GJ: ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev* 1999, 13:2339-2352.
62. Peterson CL: Chromatin remodeling enzymes: taming the machines. Third in review series on chromatin dynamics. *EMBO Rep* 2002, 3:319-322.
63. Armstrong JA, Bieker JJ, Emerson BM: A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* 1998, 95:93-104.
64. Kadam S, McAlpine GS, Phelan ML, Kingston RE, Jones KA, Emerson BM: Functional selectivity of recombinant mammalian SWI/SNF subunits. *Genes Dev* 2000, 14:2441-2451.
65. Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE: Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 1999, 98:37-46.
66. Francis NJ, Saurin AJ, Shao Z, Kingston RE: Reconstitution of a functional core polycomb repressive complex. *Mol Cell* 2001, 8:545-556.
67. Ouyang L, Chen X, Bieker JJ: 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* 1998, 273:23019-23025.
68. Crossley M, Orkin SH: Phosphorylation of the erythroid transcription factor GATA-1. *J Biol Chem* 1994, 269:16589-16596.
69. Casteel D, Suhasini M, Gudi T, Naima R, Pilz RB: Regulation of the erythroid transcription factor NF-E2 by cyclic adenosine monophosphate-dependent protein kinase. *Blood* 1998, 91:3193-3201.
70. Amrolia PJ, Ramamurthy L, Saluja D, Tanese N, Jane SM, Cunningham JM: 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* 1998, 94:10051-10056.
71. de Boer E, Rodriguez P, Bonte E, Krijgsveld J, Katsantoni E, Heck A, Grosveld F, Strouboulis J: Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* 2003, 100:7480-7485.
72. Armstrong JA, Emerson BM: NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 in vitro. *Mol Cell Biol* 1996, 16:5634-5644.
73. Bagga R, Armstrong JA, Emerson BM: Role of chromatin structure and distal enhancers in tissue-specific transcriptional regulation in vitro. *Cold Spring Harb Symp Quant Biol* 1998, 63:569-576.
74. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS: Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* 2002, 9:279-289.
75. Droge P, Muller-Hill B: High local protein concentrations at promoters: strategies in prokaryotic and eukaryotic cells. *Bioessays* 2001, 23:179-183.
76. White J, Stelzer E: Photobleaching GFP reveals protein dynamics inside live cells. *Trends Cell Biol* 1999, 9:61-65.
77. Phair RD, Misteli T: High mobility of proteins in the mammalian cell nucleus. *Nature* 2000, 404:604-609.
78. Wilson KL: The nuclear envelope, muscular dystrophy and gene expression. *Trends Cell Biol* 2000, 10:125-129.
79. Allen TD, Cronshaw JM, Bagley S, Kiseleva E, Goldberg MW: The nuclear pore complex: mediator of translocation between nucleus and cytoplasm. *J Cell Sci* 2000, 113:1651-1659.
80. Wente SR: Gatekeepers of the nucleus. *Science* 2000, 288:1374-1377.
81. Moir RD, Yoon M, Khuon S, Goldman RD: Nuclear lamins A and B1: different pathways of assembly during nuclear envelope formation in living cells. *J Cell Biol* 2000, 151:1155-1168.
82. Haraguchi T, Koujin T, Hayakawa T, Kaneda T, Tsutsumi C, Imamoto N, Akazawa C, Sukegawa J, Yoneda Y, Hiraoka Y: Live fluorescence imaging reveals early recruitment of emerin, LBR, RanBP2, and Nup153 to reforming functional nuclear envelopes. *J Cell Sci* 2000, 113:779-794.
83. Broers JL, Machiels BM, van Eys GJ, Kuijpers HJ, Manders EM, van Driel R, Ramaekers FC: Dynamics of the nuclear lamina as monitored by GFP-tagged A-type lamins. *J Cell Sci* 1999, 112:3463-3475.
84. Daigle N, Beaudouin J, Hartnell L, Imreh G, Hallberg E, Lippincott-Schwartz J, J. E: Nuclear pore complexes form immobile networks and have a very low turnover in live mammalian cells. *J Cell Biol* 2001, 154:71-84.
85. Lyman SK, Gerace L: Nuclear pore complexes: dynamics in unexpected places. *J Cell Biol* 2001, 154:17-20.
86. Li G, Sudlow G, Belmont AS: Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: precise choreography of condensation/decondensation and nuclear positioning. *J Cell Biol* 1998, 140:975-989.
87. Andrulis ED, Neiman AM, Zappulla DC, Sternglanz R: Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* 1998, 394:592-595.
88. Laroche T, Martin SG, Tsai-Pflugfelder M, Gasser SM: The dynamics of yeast telomeres and silencing proteins through the cell cycle. *J Struct Biol* 1998, 129:159-174.

89. Hediger F, Neumann FR, Van Houwe G, Dubrana K, Gasser SM: Live Imaging of Telomeres. yKu and Sir Proteins Define Redundant Telomere-Anchoring Pathways in Yeast. *Curr Biol* 2002, 12:2076-2089.
90. Gasser SM: Visualizing chromatin dynamics in interphase nuclei. *Science* 2002, 296:1412-1416.
91. Andrulis ED, Zappulla DC, Ansari A, Perrod S, Laiosa CV, Gartenberg MR, Sternglanz R: Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. *Mol Cell Biol* 2002, 22:8292-8301.
92. Gerasimova TI, Byrd K, Corces VG: A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* 2000, 6:1025-1035.
93. Marshall WF: Order and disorder in the nucleus. *Curr Biol* 2002, 12:R185-R192.
94. Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, Singh H: Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 2002, 296:158-162.
95. Bridger JM, Boyle S, Kill IR, Bickmore WA: Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. *Curr Biol* 2000, 10:149-152.
96. Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA: The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* 2001, 10:211-219.
97. Hutchison N, Weintraub H: Localization of DNAase I-sensitive sequences to specific regions of interphase nuclei. *Cell* 1985, 43:471-482.
98. Chan JK, Park PC, De Boni U: Association of DNAse sensitive chromatin domains with the nuclear periphery in 3T3 cells in vitro. *Biochem Cell Biol* 2000, 78:67-78.
99. Cremer T, Cremer C: Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2001, 2:292-301.
100. Spann TP, Goldman AE, Wang C, Huang S, Goldman RD: Alteration of nuclear lamin organization inhibits RNA polymerase II-dependent transcription. *J Cell Biol* 2002, 156:603-608.
101. Holaska JM, Wilson KL, Mansharamani M: The nuclear envelope, lamins and nuclear assembly. *Curr Opin Cell Biol* 2002, 14:357-364.
102. Dundr M, Misteli T: Functional architecture in the cell nucleus. *Biochem J* 2001, 356:297-310.
103. Alberts B, Bray D, Lewis J, Raff M, Robberts K, Watson JD: *Molecular biology of the cell*, 3rd edn. New York, NY: Garland; 1994.
104. Koberna K, Malinsky J, Pliss A, Masata M, Vecerova J, Fialova M, Bednar J, Raska I: Ribosomal genes in focus: new transcripts label the dense fibrillar components and form clusters indicative of "Christmas trees" in situ. *J Cell Biol* 2002, 157:743-748.
105. Jackson DA, Iborra FJ, Manders EM, Cook PR: Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* 1998, 9:1523-1536.
106. Jackson DA, Pombo A, Iborra F: The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. *FASEB J* 2000, 14:242-254.
107. Andersen JS, Lyon CE, Fox AH, Leung AK, Lam YW, Steen H, Mann M, Lamond AI: Directed proteomic analysis of the human nucleolus. *Curr Biol* 2002, 12:1-11.
108. Carmo-Fonseca M: The contribution of nuclear compartmentalization to gene regulation. *Cell* 2002, 108:513-521.
109. Dundr M, Hoffmann-Rohrer U, Hu Q, Grummt I, Rothblum LI, Phair RD, Misteli T: A kinetic framework for a mammalian RNA polymerase in vivo. *Science* 2002, 298:1623-1626.
110. Pombo A, Jackson DA, Hollinshead M, Wang Z, Roeder RG, Cook PR: Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *Embo J* 1999, 18:2241-2253.
111. Wijgerde M, Grosveld F, Fraser P: Transcription complex stability and chromatin dynamics in vivo. *Nature* 1995, 377:209-213.
112. Cook PR: The organization of replication and transcription. *Science* 1999, 284:1790-1795.
113. Jackson DA: The anatomy of transcription sites. *Curr Opin Cell Biol* 2003, 15:311-317.
114. Kimura H, Sugaya K, Cook PR: The transcription cycle of RNA polymerase II in living cells. *J Cell Biol* 2002, 159:777-782.
115. Harrington KS, Javed A, Drissi H, McNeil S, Lian JB, Stein JL, Van Wijnen AJ, Wang YL, Stein GS: Transcription factors RUNX1/AML1 and RUNX2/Cbfa1 dynamically associate with stationary subnuclear domains. *J Cell Sci* 2002, 115:4167-4176.
116. Pederson T: Dynamics and genome-centricity of interchromatin domains in the nucleus. *Nat Cell Biol* 2002, 4:E287-E291.
117. Misteli T, Caceres JF, Spector DL: The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 1997, 387:523-527.
118. Muratani M, Gerlich D, Janicki SM, Gebhard M, Eils R, Spector DL: Metabolic-energy-dependent movement of PML bodies within the mammalian cell nucleus. *Nat Cell Biol* 2002, 4:106-110.
119. Platani M, Goldberg I, Lamond AI, Swedlow JR: Cajal body dynamics and association with chromatin are ATP-dependent. *Nat Cell Biol* 2002, 4:502-508.
120. Fox AH, Lam YW, Leung AK, Lyon CE, Andersen J, Mann M, Lamond AI: Paraspeckles: a novel nuclear domain. *Curr Biol* 2002, 12:13-25.

121. Dillon N, Festenstein R: Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet* 2002, 18:252-258.
122. Antequera F, Bird A: Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* 1993, 90:11995-11999.
123. Cross SH, Clark VH, Simmen MW, Bickmore WA, Maroon H, Langford CF, Carter NP, Bird AP: CpG island libraries from human chromosomes 18 and 22: landmarks for novel genes. *Mamm Genome* 2000, 11:373-383.
124. Tazi J, Bird A: Alternative chromatin structure at CpG islands. *Cell* 1990, 60:909-920.
125. Delgado S, Gomez M, Bird A, Antequera F: Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J* 1998, 17:2426-2435.
126. Richards EJ, Elgin SC: Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* 2002, 108:489-500.
127. Reinhart BJ, Bartel DP: Small RNAs correspond to centromere heterochromatic repeats. *Science* 2002, 297:1831.
128. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA: Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 2002, 297:1833-1837.
129. Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI: Establishment and maintenance of a heterochromatin domain. *Science* 2002, 297:2232-2237.
130. Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP, Allshire RC: Requirement of heterochromatin for cohesion at centromeres. *Science* 2001, 294:2539-2542.
131. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T: Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 2001, 107:323-337.
132. Csink AK, Henikoff S: Genetic modification of heterochromatic association and nuclear organization in Drosophila. *Nature* 1996, 381:529-531.
133. Brown KE, Guest SS, Smale ST, Hahm K, Merkenschlager M, Fisher AG: Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* 1997, 91:845-854.
134. Brown KE, Baxter J, Graf D, Merkenschlager M, Fisher AG: Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol Cell* 1999, 3:207-217.
135. Robinett CC, Straight A, Li G, Willhelm C, Sudlow G, Murray A, Belmont AS: In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol* 1996, 135:1685-1700.
136. Heun P, Laroche T, Shimada K, Furrer P, Gasser SM: Chromosome dynamics in the yeast interphase nucleus. *Science* 2001, 294:2181-2186.
137. Chubb JR, Boyle S, Perry P, Bickmore WA: Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol* 2002, 12:439-445.
138. Vazquez J, Belmont AS, Sedat JW: Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Curr Biol* 2001, 11:1227-1239.
139. Tumbar T, Sudlow G, Belmont AS: Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol* 1999, 145:1341-1354.
140. Tsukamoto T, Hashiguchi N, Janicki SM, Tumbar T, Belmont AS, Spector DL: Visualization of gene activity in living cells. *Nat Cell Biol* 2000, 2:871-878.
141. Memedula S, Belmont AS: Sequential Recruitment of HAT and SWI/SNF Components to Condensed Chromatin by VP16. *Curr Biol* 2003, 13:241-246.
142. Kimura H, Cook PR: Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J Cell Biol* 2001, 153:1341-1353.
143. Misteli T, Gunjan A, Hock R, Bustin M, Brown DT: Dynamic binding of histone H1 to chromatin in living cells. *Nature* 2000, 408:877-881.
144. Cheutin T, McNairn AJ, Jenuwein T, Gilbert DM, Singh PB, Misteli T: Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* 2003, 299:721-725.
145. Festenstein R, Pagakis SN, Hiragami K, Lyon D, Verreault A, Sekkali B, Kioussis D: Modulation of heterochromatin protein 1 dynamics in primary Mammalian cells. *Science* 2003, 299:719-721.
146. McNally JG, Muller WG, Walker D, Wolford R, Hager GL: The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* 2000, 287:1262-1265.
147. Becker M, Baumann C, John S, Walker DA, Vigneron M, McNally JG, Hager GL: Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO Rep* 2002, 3:1188-1194.
148. Stenoien DL, Patel K, Mancini MG, Dutertre M, Smith CL, O'Malley BW, Mancini MA: FRAP reveals that mobility of oestrogen receptor-alpha is ligand- and proteasome-dependent. *Nat Cell Biol* 2001, 3:15-23.
149. Elbi C, Misteli T, Hager GL: Recruitment of dioxin receptor to active transcription sites. *Mol Biol Cell* 2002, 13:2001-2015.
150. Misteli T: Protein dynamics: implications for nuclear architecture and gene expression. *Science* 2001, 291:843-847.
151. Kamakaka RT: Heterochromatin: proteins in flux lead to stable repression. *Curr Biol* 2003, 13:R317-R319.

152. Misteli T: The concept of self-organization in cellular architecture. *J Cell Biol* 2001, 155:181-185.
153. Weatherall DJ: The Thalassemias. In: *The molecular basis of blood diseases* Edited by Stamatoyannopoulos G, Majerus P, Perlmutter R, Varmus H. pp. 183-225. Philadelphia, PA: W.B. Saunders; 2001: 183-225.
154. Hardison R: Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J Exp Biol* 1998, 201:1099-1117.
155. Antoniou M, Grosveld F: beta-globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev* 1990, 4:1007-1013.
156. Bodine DM, Ley TJ: An enhancer element lies 3' to the human A gamma globin gene. *EMBO J* 1987, 6:2997-3004.
157. Liu Q, Bungert J, Engel JD: 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* 1997, 94:169-174.
158. Cao SX, Gutman PD, Dave HP, Schechter AN: Identification of a transcriptional silencer in the 5'-flanking region of the human epsilon-globin gene. *Proc Natl Acad Sci U S A* 1989, 86:5306-5309.
159. Ronchi A, Berry M, Raguz S, Imam A, Yannoutsos N, Ottolenghi S, Grosveld F, Dillon N: Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin. *Embo J* 1996, 15:143-149.
160. Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, Miller W: Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 1997, 205:73-94.
161. Bulger M, van Doorninck JH, Saitoh N, Telling A, Farrell C, Bender MA, Felsenfeld G, Axel R, Groudine M, von Doorninck JH: Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. *Proc Natl Acad Sci U S A* 1999, 96:5129-5134.
162. Bulger M, Bender MA, van Doorninck JH, Wertman B, Farrell CM, Felsenfeld G, Groudine M, Hardison R: Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters. *Proc Natl Acad Sci U S A* 2000, 97:14560-14565.
163. Tuan D, Solomon W, Li Q, London IM: The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A* 1985, 82:6384-6388.
164. Farrell CM, Grinberg A, Huang SP, Chen D, Pichel JG, Westphal H, Felsenfeld G: A large upstream region is not necessary for gene expression or hypersensitive site formation at the mouse beta -globin locus. *Proc Natl Acad Sci U S A* 2000, 97:14554-14559.
165. Van der Ploeg LH, Konings A, Oort M, Roos D, Bernini L, Flavell RA: gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature* 1980, 283:637-642.
166. Forrester WC, Thompson C, Elder JT, Groudine M: A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A* 1986, 83:1359-1363.
167. Moon AM, Ley TJ: 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* 1990, 87:7693-7697.
168. Blom van Assendelft G, Hanscombe O, Grosveld F, Greaves DR: The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell* 1989, 56:969-977.
169. Reitman M, Lee E, Westphal H, Felsenfeld G: An enhancer/locus control region is not sufficient to open chromatin. *Mol Cell Biol* 1993, 13:3990-3998.
170. Guy LG, Kothary R, DeRepentigny Y, Delvoye N, Ellis J, Wall L: The beta-globin locus control region enhances transcription of but does not confer position-independent expression onto the lacZ gene in transgenic mice. *EMBO J* 1996, 15:3713-3721.
171. Tewari R, Gillemans N, Harper A, Wijgerde M, Zafarana G, Drabek D, Grosveld F, Philipsen S: The human beta-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene. *Development* 1996, 122:3991-3999.
172. Philipsen S, Talbot D, Fraser P, Grosveld F: The beta-globin dominant control region: hypersensitive site 2. *Embo J* 1990, 9:2159-2167.
173. Gillemans N, Tewari R, Lindeboom F, Rottier R, de Wit T, Wijgerde M, Grosveld F, Philipsen S: 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* 1998, 12:2863-2873.
174. Goodwin AJ, McInerney JM, Glander MA, Pomerantz O, Lowrey CH: In vivo formation of a human beta-globin locus control region core element requires binding sites for multiple factors including GATA-1, NF-E2, erythroid Kruppel-like factor, and Sp1. *J Biol Chem* 2001, 276:26883-26892.
175. Talbot D, Philipsen S, Fraser P, Grosveld F: Detailed analysis of the site 3 region of the human beta-globin dominant control region. *Embo J* 1990, 9:2169-2177.
176. Pruzina S, Hanscombe O, Whyatt D, Grosveld F, Philipsen S: Hypersensitive site 4 of the human beta globin locus control region. *Nucleic Acids Res* 1991, 19:1413-1419.

177. Fraser P, Pruzina S, Antoniou M, Grosveld F: Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev* 1993, 7:106-113.
178. Tuan DY, Solomon WB, London IM, Lee DP: An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A* 1989, 86:2554-2558.
179. Ney PA, Sorrentino BP, McDonagh KT, Nienhuis AW: Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev* 1990, 4:993-1006.
180. Ellis J, Tan-Un KC, Harper A, Michalovich D, Yannoutsos N, Philipsen S, Grosveld F: A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *Embo J* 1996, 15:562-568.
181. Wijgerde M, Gribnau J, Trimborn T, Nuez B, Philipsen S, Grosveld F, Fraser P: The role of EKLF in human beta-globin gene competition. *Genes Dev* 1996, 10:2894-2902.
182. Bungert J, Dave U, Lim KC, Lieuw KH, Shavit JA, Liu Q, Engel JD: Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev* 1995, 9:3083-3096.
183. Li Q, Stamatoyannopoulos JA: Position independence and proper developmental control of gamma-globin gene expression require both a 5' locus control region and a downstream sequence element. *Mol Cell Biol* 1994, 14:6087-6096.
184. Farrell CM, West AG, Felsenfeld G: Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol* 2002, 22:3820-3831.
185. Bender MA, Roach JN, Halow J, Close J, Alami R, Bouhassira EE, Groudine M, Fiering SN: Targeted deletion of 5'HS1 and 5'HS4 of the beta-globin locus control region reveals additive activity of the DNaseI hypersensitive sites. *Blood* 2001, 98:2022-2027.
186. Fiering S, Epner E, Robinson K, Zhuang Y, Telling A, Hu M, Martin DI, Enver T, Ley TJ, Groudine M: Targeted deletion of 5'HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. *Genes Dev* 1995, 9:2203-2213.
187. Hug BA, Wesselschmidt RL, Fiering S, Bender MA, Epner E, Groudine M, Ley TJ: Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Mol Cell Biol* 1996, 16:2906-2912.
188. Bender MA, Reik A, Close J, Telling A, Epner E, Fiering S, Hardison R, Groudine M: Description and targeted deletion of 5' hypersensitive site 5 and 6 of the mouse beta-globin locus control region. *Blood* 1998, 92:4394-4403.
189. Bender MA, Mehaffey MG, Telling A, Hug B, Ley TJ, Groudine M, Fiering S: Independent formation of DnaseI hypersensitive sites in the murine beta-globin locus control region. *Blood* 2000, 95:3600-3604.
190. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W: Looping and Interaction between Hypersensitive Sites in the Active beta-globin Locus. *Mol Cell* 2002, 10:1453-1465.
191. Palstra R-J, Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W.: Structural dynamics of a nuclear compartment of *cis*-regulatory DNA elements and beta-globin genes during erythroid differentiation and development. submitted.
192. Forrester WC, Epner E, Driscoll MC, Enver T, Brice M, Papayannopoulou T, Groudine M: 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* 1990, 4:1637-1649.
193. Bottardi S, Aumont A, Grosveld F, Milot E: Developmental stage-specific epigenetic control of human beta globin gene expression is set in hematopoietic progenitor cells prior to their transcriptional activation. *Blood* 2003, in press.
194. van der Ploeg LH, Flavell RA: DNA methylation in the human gamma delta beta-globin locus in erythroid and nonerythroid tissues. *Cell* 1980, 19:947-958.
195. Forsberg EC, Downs KM, Christensen HM, Im H, Nuzzi PA, Bresnick EH: Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc Natl Acad Sci U S A* 2000, 97:14494-14499.
196. Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C, Enver T: Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* 1997, 11:774-785.
197. Delassus S, Titley I, Enver T: Functional and molecular analysis of hematopoietic progenitors derived from the aorta-gonad-mesonephros region of the mouse embryo. *Blood* 1999, 94:1495-1503.
198. Schubeler D, Francastel C, Cimbora DM, Reik A, Martin DI, Groudine M: Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev* 2000, 14:940-950.
199. Strouboulis J, Dillon N, Grosveld F: Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes Dev* 1992, 6:1857-1864.
200. Ling KW, Dzierzak E: Ontogeny and genetics of the hemato/lymphopoietic system. *Curr Opin Immunol* 2002, 14:186-191.

201. Hanscombe O, Vidal M, Kaeda J, Luzzatto L, Greaves DR, Grosveld F: High-level, erythroid-specific expression of the human alpha-globin gene in transgenic mice and the production of human hemoglobin in murine erythrocytes. *Genes Dev* 1989, 3:1572-1581.
202. Peterson KR, Stamatoyannopoulos G: Role of gene order in developmental control of human gamma- and beta-globin gene expression. *Mol Cell Biol* 1993, 13:4836-4843.
203. Chada K, Magram J, Costantini F: An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature* 1986, 319:685-689.
204. Raich N, Enver T, Nakamoto B, Josephson B, Papayannopoulou T, Stamatoyannopoulos G: Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science* 1990, 250:1147-1149.
205. Dillon N, Grosveld F: Human gamma-globin genes silenced independently of other genes in the beta-globin locus. *Nature* 1991, 350:252-254.
206. Raich N, Papayannopoulou T, Stamatoyannopoulos G, Enver T: Demonstration of a human epsilon-globin gene silencer with studies in transgenic mice. *Blood* 1992, 79:861-864.
207. Shih DM, Wall RJ, Shapiro SG: A 5' control region of the human epsilon-globin gene is sufficient for embryonic specificity in transgenic mice. *J Biol Chem* 1993, 268:3066-3071.
208. Tanimoto K, Liu Q, Grosveld F, Bungert J, Engel JD: Context-dependent EKLF responsiveness defines the developmental specificity of the human epsilon-globin gene in erythroid cells of YAC transgenic mice. *Genes Dev* 2000, 14:2778-2794.
209. Tanabe O, Katsuoka F, Campbell AD, Song W, Yamamoto M, Tanimoto K, Engel JD: An embryonic/fetal beta-type globin gene repressor contains a nuclear receptor TR2/TR4 heterodimer. *Embo J* 2002, 21:3434-3442.
210. Giglioni B, Casini C, Mantovani R, Merli S, Comi P, Ottolenghi S, Saglio G, Camaschella C, Mazza U: A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and beta-thalassemia. *EMBO J* 1984, 3:2641-2645.
211. de Villiers J, Schaffner W: A small segment of polyoma virus DNA enhances the expression of a cloned beta-globin gene over a distance of 1400 base pairs. *Nucleic Acids Res* 1981, 9:6251-6264.
212. Enver T, Raich N, Ebens AJ, Papayannopoulou T, Costantini F, Stamatoyannopoulos G: Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* 1990, 344:309-313.
213. Hanscombe O, Whyatt D, Fraser P, Yannoutsos N, Greaves D, Dillon N, Grosveld F: Importance of globin gene order for correct developmental expression. *Genes Dev* 1991, 5:1387-1394.
214. Dillon N, Trimborn T, Strouboulis J, Fraser P, Grosveld F: The effect of distance on long-range chromatin interactions. *Mol Cell* 1997, 1:131-139.
215. Tanimoto K, Liu Q, Bungert J, Engel JD: Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature* 1999, 398:344-348.
216. Gribnau J, de Boer E, Trimborn T, Wijgerde M, Milot E, Grosveld F, Fraser P: Chromatin interaction mechanism of transcriptional control in vivo. *Embo J* 1998, 17:6020-6027.
217. Trimborn T, Gribnau J, Grosveld F, Fraser P: Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci. *Genes Dev* 1999, 13:112-124.
218. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P: Long-range chromatin regulatory interactions in vivo. *Nat Genet* 2002, 32:623-626.
219. Ptashne M: Gene regulation by proteins acting nearby and at a distance. *Nature* 1986, 322:697-701.
220. Grosveld F: Activation by locus control regions? *Curr Opin Genet Dev* 1999, 9:152-157.
221. de Krom M, van de Corput M, von Lindern M, Grosveld F, Strouboulis J: Stochastic patterns in globin gene expression are established prior to transcriptional activation and are clonally inherited. *Mol Cell* 2002, 9:1319-1326.
222. Becker PB: Nucleosome sliding: facts and fiction. *EMBO J* 2002, 21:4749-4753.
223. Tuan D, Kong S, Hu K: Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A* 1992, 89:11219-11223.
224. Herendeen DR, Kassavetis GA, Geiduschek EP: A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science* 1992, 256:1298-1303.
225. Dorsett D: Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr Opin Genet Dev* 1999, 9:505-514.
226. Bulger M, Groudine M: Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* 1999, 13:2465-2477.
227. Bulger M, Groudine M: TRAPping enhancer function. *Nat Genet* 2002, 32:555-556.
228. Rippe K: Making contacts on a nucleic acid polymer. *Trends Biochem Sci* 2001, 26:733-740.
229. Dekker J, Rippe K, Dekker M, Kleckner N: Capturing chromosome conformation. *Science* 2002, 295:1306-1311.
230. Splinter E, Grosveld, F., and de Laat, W.: 3C technology: analyzing the spatial organisation of genomic loci in vivo. submitted.
231. Talbot D, Collis P, Antoniou M, Vidal M, Grosveld F, Greaves DR: A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* 1989, 338:352-355.

232. Imam AM, Patrinos GP, de Krom M, Bottardi S, Janssens RJ, Katsantoni E, Wai AW, Sherratt DJ, Grosveld FG: Modification of human beta-globin locus PAC clones by homologous recombination in Escherichia coli. *Nucleic Acids Res* 2000, 28:E65.
233. de Krom M: *Activation, regulation and transcription of the human and murine globin loci*. Rotterdam: Erasmus MC; 2002.
234. Scrable H, Stambrook PJ: Activation of the lac repressor in the transgenic mouse. *Genetics* 1997, 147:297-304.
235. Wyborski DL, DuCoeur LC, Short JM: Parameters affecting the use of the lac repressor system in eukaryotic cells and transgenic animals. *Environ Mol Mutagen* 1996, 28:447-458.
236. Cronin CA, Gluba W, Scrable H: The lac operator-repressor system is functional in the mouse. *Genes Dev* 2001, 15:1506-1517.
237. Lewandoski M: Conditional control of gene expression in the mouse. *Nat Rev Genet* 2001, 2:743-755.
238. von Lindern M, Parren-van Amelsvoort M, van Dijk T, Deiner E, van den Akker E, van Emst-de Vries S, Willems P, Beug H, Lowenberg B: Protein kinase C alpha controls erythropoietin receptor signaling. *J Biol Chem* 2000, 275:34719-34727.
239. von Lindern M, Zauner W, Mellitzer G, Steinlein P, Fritsch G, Huber K, Lowenberg B, Beug H: The glucocorticoid receptor cooperates with the erythropoietin receptor and c-Kit to enhance and sustain proliferation of erythroid progenitors in vitro. *Blood* 1999, 94:550-559.
240. Pear WS, Nolan GP, Scott ML, Baltimore D: Production of high-titer helper-free retroviruses by transient transfection. *Proc Natl Acad Sci U S A* 1993, 90:8392-8396.
241. Whitaker M: Fluorescent tags of protein function in living cells. *Bioessays* 2000, 22:180-187.
242. Day RN, Periasamy A, Schaufele F: Fluorescence resonance energy transfer microscopy of localized protein interactions in the living cell nucleus. *Methods* 2001, 25:4-18.
243. Gossen M, Bujard H: Tight control of gene expression in mammalian cells by tetracycline- responsive promoters. *Proc Natl Acad Sci U S A* 1992, 89:5547-5551.
244. Antoniou M: Introduction of erythroid-specific expression in murine erythroidleukemia (MEL) cell lines. In: *Methods in Molecular Biology* Edited by Murray EJ, vol. 7. Clifton, NJ: The Humana Press Inc; 1991.
245. Dillon N, Grosveld F: Transcriptional regulation of multigene loci: multilevel control. *Trends Genet* 1993, 9:134-137.
246. Fraser P, Hurst J, Collis P, Grosveld F: DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res* 1990, 18:3503-3508.
247. Wall L, deBoer E, Grosveld F: The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev* 1988, 2:1089-1100.
248. Wasylyk B, Wasylyk C, Augereau P, Chambon P: The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. *Cell* 1983, 32:503-514.
249. Banerji J, Rusconi S, Schaffner W: Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 1981, 27:299-308.
250. Gerasimova TI, Corces VG: Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu Rev Genet* 2001, 35:193-208.
251. Muller HP, Schaffner W: Transcriptional enhancers can act in trans. *Trends Genet* 1990, 6:300-304.
252. Bickel S, Pirrotta V: Self-association of the Drosophila zeste protein is responsible for transvection effects. *EMBO J* 1990, 9:2959-2967.
253. Mueller-Storm HP, Sogo JM, Schaffner W: An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell* 1989, 58:767-777.
254. Dunaway M, Droge P: Transactivation of the Xenopus rRNA gene promoter by its enhancer. *Nature* 1989, 341:657-659.
255. Mahmoudi T, Katsani KR, Verrijzer CP: GAGA can mediate enhancer function in trans by linking two separate DNA molecules. *EMBO J* 2002, 21:1775-1781.
256. Heuchel R, Matthias P, Schaffner W: Two closely spaced promoters are equally activated by a remote enhancer: evidence against a scanning model for enhancer action. *Nucleic Acids Res* 1989, 17:8931-8947.
257. de Bruin D, Zaman Z, Liberatore RA, Ptashne M: Telomere looping permits gene activation by a downstream UAS in yeast. *Nature* 2001, 409:109-113.
258. Lundgren M, Chow CM, Sabbattini P, Georgiou A, Minaee S, Dillon N: Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* 2000, 103:733-743.
259. Locke J, Kotarski MA, Tartof KD: Dosage-dependent modifiers of position effect variegation in Drosophila and a mass action model that explains their effect. *Genetics* 1988, 120:181-198.
260. Burgess-Beusse B, Farrell C, Gaszner M, Litt M, Mutskov V, Recillas-Targa F, Simpson M, West A, Felsenfeld G: The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A* 2002, 99 Suppl 4:16433-16437.
261. Bulger M, Sawado T, Schubeler D, Groudine M: ChIPs of the beta-globin locus: unraveling gene regulation within an active domain. *Curr Opin Genet Dev* 2002, 12:170-177.

262. Cremer T, Kreth G, Koester H, Fink RH, Heintzmann R, Cremer M, Solovei I, Zink D, Cremer C: Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit Rev Eukaryot Gene Expr* 2000, 10:179-212.
263. Sabbattini P, Georgiou A, Sinclair C, Dillon N: Analysis of mice with single and multiple copies of transgenes reveals a novel arrangement for the lambda5-VpreB1 locus control region. *Mol Cell Biol* 1999, 19:671-679.
264. Groudine M, Weintraub H: Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. *Cell* 1982, 30:131-139.
265. Morcillo P, Rosen C, Baylies MK, Dorsett D: Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. *Genes Dev* 1997, 11:2729-2740.
266. Rollins RA, Morcillo P, Dorsett D: Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* 1999, 152:577-593.
267. Zhou J, Ashe H, Burks C, Levine M: Characterization of the transvection mediating region of the abdominal-B locus in Drosophila. *Development* 1999, 126:3057-3065.
268. Zhou J, Levine M: A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the Drosophila embryo. *Cell* 1999, 99:567-575.
269. Sipos L, Mihaly J, Karch F, Schedl P, Gausz J, Gyurkovics H: Transvection in the Drosophila Abd-B domain: extensive upstream sequences are involved in anchoring distant cis-regulatory regions to the promoter. *Genetics* 1998, 149:1031-1050.
270. Calzolari R, McMorrow T, Yannoutsos N, Langeveld A, Grosveld F: 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* 1999, 18:949-958.
271. Kmita M, Kondo T, Duboule D: Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. *Nat Genet* 2000, 26:451-454.
272. Isogai Y, Tjian R: Targeting genes and transcription factors to segregated nuclear compartments. *Curr Opin Cell Biol* 2003, 15:296-303.
273. Chubb JR, Bickmore WA: Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell* 2003, 112:403-406.
274. Hardison R, Miller W: Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol Biol Evol* 1993, 10:73-102.
275. Jackson DA, Hassan AB, Errington RJ, Cook PR: Visualization of focal sites of transcription within human nuclei. *Embo J* 1993, 12:1059-1065.
276. Donze D, Townes TM, Bieker JJ: Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching. *J Biol Chem* 1995, 270:1955-1959.
277. Ben-David Y, Bani MR, Chabot B, De Koven A, Bernstein A: Retroviral insertions downstream of the heterogeneous nuclear ribonucleoprotein A1 gene in erythroleukemia cells: evidence that A1 is not essential for cell growth. *Mol Cell Biol* 1992, 12:4449-4455.
278. Lu SJ, Rowan S, Bani MR, Ben-David Y: Retroviral integration within the Fli-2 locus results in inactivation of the erythroid transcription factor NF-E2 in Friend erythroleukemias: evidence that NF-E2 is essential for globin expression. *Proc Natl Acad Sci U S A* 1994, 91:8398-8402.
279. Dolznig H, Boulme F, Stangl K, Deiner EM, Mikulits W, Beug H, Mullner EW: Establishment of normal, terminally differentiating mouse erythroid progenitors: molecular characterization by cDNA arrays. *FASEB J* 2001, 15:1442-1444.
280. von Lindern M, Deiner EM, Dolznig H, Parren-Van Amelsvoort M, Hayman MJ, Mullner EW, Beug H: Leukemic transformation of normal murine erythroid progenitors: v- and c-ErbB act through signaling pathways activated by the EpoR and c-Kit in stress erythropoiesis. *Oncogene* 2001, 20:3651-3664.
281. Aihara T, Yasuo M, Kumiko,, K. S, Y., , Imaoka S, Monden M, Nakamura Y: Genes preferentially expressed in precancerous lesion of HCC. *(unpublished)* 1998.
282. Wadman IA, Osada H, Grutz GG, Agulnick AD, Westphal H, Forster A, Rabbitts TH: The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J* 1997, 16:3145-3157.
283. Lecuyer E, Herblot S, Saint-Denis M, Martin R, Begley CG, Porcher C, Orkin SH, Hoang T: The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. *Blood* 2002, 100:2430-2440.
284. McMorrow T, van den Wijngaard A, Wollenschlaeger A, van de Corput M, Monkhorst K, Trimborn T, Fraser P, van Lohuizen M, Jenuwein T, Djabali M, Philipsen S, Grosveld F, Milot E: Activation of the beta globin locus by transcription factors and chromatin modifiers. *Embo J* 2000, 19:4986-4996.
285. Philipsen S, Pruzina S, Grosveld F: The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. *Embo J* 1993, 12:1077-1085.
286. Bi X, Broach JR: Chromosomal boundaries in S. cerevisiae. *Curr Opin Genet Dev* 2001, 11:199-204.

287. Vyas P, Vickers MA, Simmons DL, Ayyub H, Craddock CF, Higgs DR: Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. *Cell* 1992, 69:781-793.
288. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ: Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 1997, 389:251-260.
289. Horn PJ, Peterson CL: Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science* 2002, 297:1824-1827.
290. Aboussekhra A, Biggerstaff M, Shivji MKK, Vilpo JA, Moncollin V, Podust VN, Protic M, Hubscher U, Egly J-M, Eood RD: Mammalian DNA nucleotide excision repair reconstituted with purified components. *Cell* 1995, 80:859-868.
291. Evans E, Fellows J, Coffer A, Wood R: Open complex formation around a lesion during nucleotide excision repair factors. *EMBO J* 1997, 16:625-638.
292. Evans E, Moggs J, Hwang J, Egly J, Wood R: Mechanism of open complex and dual incision formation by human nucleotide exision repair factors. *EMBO J* 1997, 16:6559-6573.

Summary

RNA polymerase II mainly transcribes protein-encoding genes, and the transcriptional activity of many of those genes is tightly regulated. Two cooperating components are key in regulating RNAP II transcription. One component acts *in cis* and requires an element that is located on the same DNA molecule as the transcribed gene, the so-called *cis*-regulatory DNA elements. The other component acts *in trans*, i.e. it requires molecules separate from the gene containing DNA molecule, here referred to as *trans*-acting factors.

Despite the fact that *cis*-regulatory DNA elements have different functional definitions, there are two features that they have in common: they can be detected as DNase I hypersensitive sites (HSs) *in vivo* and they are relatively small DNA fragments (200-300bp) that contain multiple binding sites for certain *trans*-acting factors and their interacting partners. *Trans*-acting factors are a diverse group of proteins (and RNA), which includes sequence-specific transcription factors, components of the general transcription machinery, and co-factors. Many *trans*-acting factors can interact and as such influence each other activities.

How do *cis*-regulatory DNA elements and *trans*-acting factors collaborate efficiently to regulate RNAP II transcription? Furthermore, *cis*-regulatory DNA elements somehow need to communicate with each other. However, many gene loci have widely separated regulatory elements that all appear to contribute to proper transcriptional regulation. One important aspect of (bio)chemistry is that the reaction rate is proportional to the concentration of molecules involved. An economical way to bring the different components together might be through compartmentalization.

Indeed, the nucleus of eukaryotes, which holds the genetic material (DNA), is organized into many distinct compartments each dedicated to drive particular biochemical processes. The DNA of the nucleus is packaged into chromatin, which has constrained mobility and consequently a particular gene and its regulatory sequences are limited to a tiny portion of the nucleus. In contrast, *trans*-acting factors diffuse 'freely' and can roam the entire nuclear space. A rapid flux of proteins on/off macromolecules (e.g. chromatin) and in/out of compartments generates a dynamic but stable nuclear organization.

In this thesis the β-globin loci of man and mouse were used as model systems to study their transcriptional regulation in the context of the living nucleus. These loci consist of several genes and *cis*-regulatory DNA elements. Transcription of the genes is regulated in a tissue- and developmental stage-specific manner. One of the *cis*-regulatory DNA elements is the Locus Control Region (LCR), which is thought to create an open chromatin conformation and activate transcription of the genes. A hypothesis that explains this activity is the looping model. This model presumes that the LCR directly and stably interacts with individual genes, thereby activating it. The intervening chromatin between LCR and the gene is thought to loop out.

In chapter 3, this model was tested on the mouse β-globin locus *in vivo*, using a novel technique called Chromosome Conformation Capture (3C). The results showed that in erythroid cells the LCR and actively transcribed genes were in close spatial proximity, although they are separated by 40-60 kb of intervening DNA. Inactive genes located between the LCR and active genes did not interact, but looped out. In nonerythroid cells, which do not transcribe globin genes, no spatial interactions were found between the LCR and any of the genes. Interestingly, there were additional *cis*-regulatory DNA elements that joined the LCR-active gene interactions. The upstream 5' HS-62.5/-60.7 participate in this interaction, again with the intervening DNA looping out. At the other end of the locus the 3' HS1 is also involved in the contacts, but we have no evidence for DNA looping out between the genes and 3' HS1. We propose that clustering of *cis*-regulatory DNA elements is essential for creating and maintaining active chromatin domains and regulating transcription. We refer to this clustering as the 'active chromatin hub' (ACH).

In chapter 4, we examined whether the ACH is formed at other developmental stages and if its formation is evolutionary conserved between the mouse and human loci. The 3C results showed that during different stages of development an erythroid-specific spatial cluster of *cis*-regulatory elements and active genes is formed in the β-globin cluster. The core of this spatial cluster consists of the locus control region (LCR) and distal regulatory elements, a so-called core ACH. This core ACH is structurally conserved during development, while a switch occurs in globin genes interacting with the core ACH, correlating with the switch in their transcriptional activity. This spatial organisation of the β-globin gene cluster is conserved between mouse and man. Furhtermore, we found in erythroid progenitor cells that the locus adopts a conformation that differs from brain and erythroid cells. The distal hypersensitive regions and the 5' side of the LCR interact, while the inactive genes and the 5' HS1-3 of the LCR seem to loop out. These data strongly suggest that the *cis*-regulatory DNA elements surrounding the β-globin genes create an erythroid-specific developmentally stable nuclear compartment dedicated to RNA polymerase II transcription. This spatial clustering of regulatory elements results in a high local concentration of DNA binding sites for cognate transcription factors and their interacting partners, which consequently accumulate at the site. In chapter 5, we discuss several speculative properties of the ACH that may explain its role in efficient transcription of the globin genes.

Globin genes are transcribed alternately rather than at the same time and transcription can switch back and forth between genes in a flip-flop mechanism. This implies that interactions between the different components of the ACH are a dynamic rather than static process. However, the 3C measurements represent steady-state average levels. Consequently, the dynamic interactions cannot be detected. In chapter 2, we describe an approach to monitor these dynamic interactions in living cells. We 'marked' the human β-globin locus with a dual tagging system, using Lac^{O}/ LacR-cyan fluorescent protein (CFP) in conjunction with a Tet^{O}/TetR-yellow fluorescent protein (YFP). This work is still in progress.

Samenvatting

Het onderzoek waar ik de afgelopen jaren bij betrokken ben geweest, richt zich op de β-globine genen en in het bijzonder op de vraag hoe de activiteit van deze genen gereguleerd wordt in de celkern van bloedcellen. Allereerst zal ik een korte inleiding geven, vervolgens de doelstelling van het onderzoek en tot slot de behaalde resultaten beschrijven.

De term actief gen betekent dat er transcriptie plaats vindt op dat gen. Transcriptie is het proces waarbij een enkele streng van het genetische materiaal (d.w.z. het DNA) wordt gekopieerd naar een complementaire RNA molecuul. Dit proces wordt uitgevoerd door speciale eiwitcomplexen genaamd RNA polymerases (RNAPs) en de celkern bevat drie verschillende RNAPs. Eén daarvan, RNAP II, is voornamelijk betrokken bij de transcriptie van genen die coderen voor eiwitten, zoals de β-globine genen. Transcriptie is niet een spontaan proces, maar gebeurt onder strikte regulatie in de celkern. Allereerst, begint transcriptie op een specifieke plek op het DNA ten opzichte van het gen om een functioneel RNA product te krijgen. Dit zijn de zogenaamde promoter sequenties die slechts enkele honderden basenparen (d.w.z. DNA bouwstenen) groot zijn, terwijl het gehele DNA in de kern (het genoom) enkele miljarden basenparen beslaat. RNAP II moet als het ware deze kleine promoter elementen herkennen om er vervolgens te “landen” en transcriptie te starten. Ten tweede, elke celkern bevat het complete genoom, maar niet alle genen (een mens heeft ongeveer 30 duizend genen) zijn actief. Tijdens de ontwikkeling van een organisme, zoals mens of muis, wordt bepaald welke genen in welke cellen actief zijn. Bovendien kunnen bepaalde celtypen de expressie hun genen nog veranderen tijdens het volwassen leven. Bijvoorbeeld: de vorming van bloedcellen gebeurt continu, hierbij wordt een zogenaamde stamcel via verschillende celdelingen omgevormd tot een bepaald type bloedcel. Dit gaat gepaard met het activeren en inactiveren van bepaalde genen. Tot slot, het DNA in een celkern is “verpakt” in bepaalde eiwitten en gezamenlijk vormen zij het chromatine (zie figuur 1.2.1). Dankzij deze verpakking zijn RNAP complexen niet in staat zijn transcriptie uit te voeren. Deze drie omstandigheden zorgen ervoor dat RNAP II complexen hulp nodig hebben van andere componenten die promoter elementen herkennen, bepalen wanneer een gen actief is, en de chromatine structuur beïnvloeden, om uiteindelijk transcriptie mogelijk te maken.

Die componenten zijn op te delen in twee categorieën. Ten eerste zijn er componenten die functioneren *in cis*, d.w.z. het zijn DNA elementen die op hetzelfde molecuul gelegen zijn als het gen dat transcriptie ondergaat. Dit zijn de zogenaamde *cis*-regulerende DNA elementen. Deze elementen bestaan naast de al eerder genoemde promoter uit enkele andere elementen allen met verschillende functionele definities. Echter, deze elementen lijken twee eigenschappen gemeen te hebben. Het zijn korte DNA sequenties (enkele honderden basenparen) waarin veel unieke korte sequenties (van enige basenparen groot) voorkomen. Daarnaast is de structuur van hun chromatine is wezenlijk anders dan het omliggende chromatine. Het lijkt meer toegankelijk te zijn voor eiwitten. Naast de *cis*-regulerende DNA elementen zijn er componenten die functioneren *in trans*, d.w.z. het zijn voornamelijk eiwit (maar ook RNA) moleculen die niet deel uitmaken van het gen bevattende DNA molecuul. Hier omschreven als *trans*-werkende factoren. Deze factoren behelzen eiwitten die deel uitmaken van de zogenaamde ‘transcriptie machine’ (waaronder RNAP II), factoren die chromatine structuur kunnen beïnvloeden, en transcriptie factoren die de korte unieke sequenties in *cis*-regulerende DNA elementen kunnen herkennen en binden. Al deze factoren kunnen binden aan elkaar en deze interacties hebben invloed op transcriptie van genen. De samenstelling en concentratie van deze *trans*-werkende factoren kunnen per celtype verschillen. Zo heeft een bloedcel een andere samenstelling van factoren dan een hersencel en daarom zijn er verschillende genen actief in de beide celtypen. Bij veel genen is waargenomen dat een juiste samenstelling van *cis*-regulerende DNA

elementen en *trans*-werkende factoren noodzakelijk is om transcriptie te krijgen. De β-globine genen vormen hierop geen uitzondering.

RNA transcripten van β-globine genen worden omgezet in het eiwit β-globine. Samen met α-globine vormt β-globine het eiwitcomplex hemoglobine dat er voor zorgt dat zuurstof van de longen naar de verschillende organen in ons lichaam wordt vervoerd. Indien er iets fout gaat bij het maken van het globine eiwit onstaat er een bloedcel die deze taak niet naar behoren kan uitvoeren. Er zijn twee groepen patiënten met zulke bloedcellen: patiënten met een niet-functioneel globine eiwit en patiënten die helemaal geen globine eiwitten maken. In beide gevallen is aangetoond dat dit wordt veroorzaakt door foutjes in het DNA. Dit heeft geleidt tot veel onderzoek naar de regulatie van β-globine transcriptie.

De β-globine genen zijn gelegen in een cluster (een locus). De mens heeft vijf globine genen en de muis vier (weergegeven als rode blokjes in figuur 1.4.2) en elk gen heeft zijn eigen promoter. Tevens bevat zowel het menselijk- als het muizenlocus enkele andere *cis*-regulerende DNA elementen (pijlen in figuur 1.4.2). Ook is er een heel scala aan *trans*-werkende factoren betrokken bij de transcriptie van deze genen. De genen zijn alleen actief in rode bloedcellen en tijdens de ontwikkeling verandert het expressie patroon van de genen. Voor de mens houdt dit in dat tijdens de eerste acht weken van de embryonale ontwikkeling het ε-globine gen actief is. Vervolgens zijn tot aan de geboorte de genen $^{G}\gamma$ en $^{A}\gamma$ actief. Na de geboorte worden de δ en β genen geactiveerd. Een aantal jaren geleden is aangetoond dat de globine genen naast de promoter ook afhankelijk zijn van enkele *cis*-regulerende DNA elementen die relatief ver van de genen liggen. Deze combinatie van *cis*-regulerende DNA elementen wordt ook wel de “Locus Control Region” (LCR) genoemd. Een vraag die hierbij ontstond was hoe relatief ver uit elkaar gelegen *cis*-regulerende DNA elementen transcriptie kunnen beïnvloeden? Bovendien suggereert het dat de LCR en de β-globine genen op een of andere manier met elkaar communiceren.

De doelstelling van het onderzoek waar ik bij betrokken ben geweest was het verkrijgen van meer inzicht in de regulatie van de β-globine genen. Hierbij werd er voornamelijk geconcentreerd op de manier van communicatie tussen de LCR en de genen. Bij het onderzoek gingen we uit van een model dat verklaarde hoe de LCR de verschillende genen van het β-globine locus zou kunnen activeren. In dit model functioneren de afzonderlijk *cis*-regulerende DNA elementen van de LCR als een eenheid, een “holocomplex”, dat bijeen gehouden wordt door *trans*-werkende factoren. Het “holocomplex” gaat een directe interactie aan met de *trans*-werkende factoren die gebonden zijn aan een promoter van een globine gen. Door deze interactie wordt dit gen geactiveerd. Het DNA dat tussen de LCR en het gen ligt vormt dan als het ware een lus en heeft geen functie. Communicatie tussen LCR en genen is in dit model gebaseerd op een directe interactie tussen beiden (zie figuur 1.4.5). Het onderzoek beschreven in dit proefschrift beproeft dit model aan de hand van experimenten met het menselijke β-globine locus in muizen, maar ook door studie van het β-globine locus van de muis zelf.

De experimenten beschreven in hoofdstuk 3 meten de 3-dimensionale structuur van het muizen β-globine locus in de celkern van hersencellen en rode bloedcellen. De resultaten laten zien dat het locus in hersencellen, waar β-globine genen niet actief zijn, een lineaire structuur aanneemt. In de kern van een rode bloedcel bevinden LCR elementen en twee actieve β-globine genen zich dicht bij elkaar in de 3-dimensionale ruimte. Ofschoon de LCR en deze genen gescheiden worden door ongeveer 40 tot 60 duizend basenparen. De inactieve β-globine genen, die gelegen zijn tussen de LCR en de actieve genen, zijn niet dichtbij gelokaliseerd, maar maken deel uit van een lus. Deze resultaten leveren direct bewijs voor het model waarbij communicatie tussen LCR en genen is

gebaseerd op een directe interactie tussen beiden. Bovendien zijn er twee andere *cis*-regulerende DNA elementen die ook dichtbij de LCR en actieve genen gelegen zijn. Deze elementen zijn aan de buitenzijden van het locus gelegen. De laatste vinding wijst er op dat *cis*-regulerende DNA elementen een cluster vormen in een locus waar actieve transcriptie plaatsvindt. Deze cluster vorming hebben we “Active Chromatin Hub” (ACH) genoemd (zie figuur 3.7).

In hoofdstuk 4 wordt aangetoond dat niet alleen het muizen β-globine locus een ACH vormt, maar ook het menselijke locus gemeten in transgene muizen. De ACH wordt gevormd tijdens het differentiatie proces van rode bloedcellen. In de voorlopers van de rode bloedcellen vindt geen transcriptie plaats, maar er bestaat al wel een structuur in het locus. Het lijkt er op dan de *cis*-regulerende elementen aan de buitenkanten van het locus en de voorkant van de LCR met elkaar interacteren (zie figuur 4.5). Tevens nemen we waar dat tijdens verschillende stadia van de ontwikkeling er een soort kernstructuur in de ACH aanwezig is. Deze kernstructuur is een interactie tussen alle elementen van de LCR en de *cis*-regulerende DNA elementen aan de buitenzijden van het locus. De β-globine genen gaan alleen een directe interactie met de kernstructuur aan wanneer zij transcriptioneel actief zijn. De kernstructuur wordt alleen gevormd in rode bloedcellen en is stabiel tijdens de ontwikkeling. De ruimtelijke interactie tussen de kernstructuur en een gen leidt tot een lokaal verhoogde concentratie van consensus-sequenties waar transcriptie factoren kunnen binden. Uiteindelijk leidt dit tot een ophoping van *trans*-werkende factoren binnen de ACH structuur (zie figuur 4.5). Lokale ophoping (of hoge concentratie) van eiwitten is een fenomeen wat vaker wordt waargenomen in de ingewikkelde structuur van een celkern en wordt compartiment vorming genoemd. Een van die compartimenten, de nucleolus, is noodzakelijk voor transcriptie van de zogenaamde ribosmale RNA genen. Compartiment vorming leidt in dit geval tot efficiënte transcriptie. Dit kan worden verklaard uit het feit dat de reactiesnelheid van een (bio)chemische reactie afhankelijk is van de concentratie van de betrokken moleculen. In hoofdstuk 5 worden de mogelijke gevolgen van een ACH compartiment vorming bediscussieerd.

Interessant is wel dat de meeste compartimenten in de celkern in eerste instantie altijd microscopisch werden waargenomen. De ACH is echter te klein om microscopisch waar te nemen. Hoewel in hoofdstuk 2 een techniek beschreven wordt die het mogelijk moet maken om een menselijk β-globine locus in levende celkernen te volgen met een microscoop. Dit werk is nog niet af, maar er wordt nog aan gewerkt om het mogelijk te maken. Mocht dit uiteindelijk lukken dan heeft deze techniek nog wat toe te voegen aan de techniek beschreven in hoofdstukken 3 en 4. Eerder is namelijk aangetoond dat transcriptie van de β-globine genen één voor één plaatsvindt en dat transcriptie tussen verschillende actieve genen voortdurend afwisselt. Dit toont aan dat transcriptie van de genen een dynamisch proces is. Echter, de techniek beschreven in hoofdstukken 3 en 4 is niet in staat om deze dynamiek waar te nemen. Het volgen van de verschillende *cis*-regulerende DNA elementen in levende celkernen zou dit wel kunnen. Dit kan waardevolle informatie opleveren over regulatie van transcriptie, onder andere door te testen hoe stabiel de interacties in de ACH zijn en welke *trans*-werkende factoren essentieel zijn voor stabiliteit.

Abbreviations

ACH	Active Chromatin Hub
AGM	Aorta/gonad/mesonephros
ATP	Adenosine triphosphate
bp	Base pairs
3C	Chromosome Conformation Capture
CFP	Cyan fluorescent protein
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
DRED	Direct repeat erythroid-definitive
EKLF	Erythroid kruppel like factor
FISH	Fluorescence *in situ* hybridization
FLIP	Fluorescence loss in photobleaching
FRAP	Fluorescence recovery after photobleaching
GFP	Green fluorescent protein
GTF	General transcription factor
HAT	Histone acetyl transferase
HDAC	Histone deacetylases
HMTase	Histone methyl transferase
HPFH	Hereditary persistence of fetal hemoglobin
HS	Hypersensitive site
kb	Kilo base pairs (i.e. 10^3 bp)
Lac^O	Lac operator
LacR	Lac repressor
LCR	Locus Control Region
MEL	Mouse erythroleukemia
mRNA	Messenger RNA
NE	Nuclear envelope
NPC	Nuclear pore complex
OR	Olfactory receptor
PAC	P1 artificial chromosome
PEV	Position effect variegation
PIC	Pre-initiation complex
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNA TRAP	RNA tagging and recovery of associated proteins
rRNA	Ribosomal RNA
TAF	TBP-associated factors
TBP	TATA-binding protein
Tet^O	Tet operator
TetR	Tet repressor
TFIIx	Transcription factor II x (e.g. B, D, E, F, H)
YFP	Yellow fluorescent protein

Curriculum Vitae

Naam:	Bastiaan Johannes Tolhuis
Geboren:	5 december 1974 te Alkmaar
Juni 1994	Eindexamen VWO, RSG Wiringherlant
September 1994- December 1995	Propedeuse Biologie, faculteit der Wiskunde en Natuurwetenschappen, Rijksuniversiteit Groningen.
December 1995- December 1998	Doctoraal Biologie, faculteit Biologie, Universiteit van Amsterdam
December 1998- September 2003	Promotieonderzoek, Erasmus MC Rotterdam, afdeling Celbiologie en Genetica. promotor: Prof. Dr. F.G. Grosveld co-promotor: Dr. W.L. de Laat
September 2003- heden	Post-doc, Nederlands Kanker Instituut, dept. Molecular Genetics

Dankwoord

Frank en Wouter veel dank voor al het geduld en vertrouwen. Frank ik heb niet echt de deur bij je plat gelopen, maar als ik dan langskwam was je altijd bereidt te luisteren en advies te geven. Ook de laatste weken bij het afronden van dit proefschrift was je hulp van grote waarde. Wouter pas halverwege mijn AIO periode raakte je betrokken bij het “GFP-verhaal”. Mede dankzij jou bleven we de problemen structureel en systematisch aanpakken, en dat moet als een hele prestatie gezien worden als je werkt met een chaoot zoals ik. Ook lukte je het telkens weer om mij te motiveren of even de gedachten op andere zaken te richten dan het “autistische wetenschappelijk onderzoek”. De maaltijd bij de Indier in Londen zal ik niet snel vergeten. Na een korte brainstorm sessie had je “de controle” voor het 3C-experiment bedacht. De weken van buffelen en pipetteren die volgden waren wat mij betreft een geweldige ervaring.

Erik en Robert-Jan het laatste jaar hebben we ons helemaal “blind ge-3C-ed”, maar het was wat mij betreft een voorrecht om dit samen met jullie te mogen doen. Het was gezellig, spannend en ook leerzaam. Ik zal deze samenwerking gaan missen in Amsterdam. Erik je hebt “gouden handjes”, maar dat wist jij natuurlijk al (hoewel het goed is als andere mensen dat ook eens lezen). Je zult er vanaf september een hoop ruimte bij krijgen aan “ons” bureau. Robert-Jan het was een waar genoegen om samen te pipetteren, drinken en discussiëren. De zomer van 2002 werd voor mij een feest door het gezamenlijk pipetteren onder het genot van een goed stukje muziek (niet waar Wouter?) Heren jullie zijn ook mijn paranimfen, alvast heel veel dank daarvoor, en ik wens jullie het allerbeste toe.

Lab 702, het waren fijne jaren. Ernie, Arnie, Mariette, Miyata, George, John Kong-a-San, en de rest heel veel dank voor al jullie hulp en gezelligheid.

Arthur en Joost bench-maatjes van de eerste jaren jullie ervaring was heel leerzaam en nuttig. Arthur je kunt als een echte Leidenaar vreselijk veel “ouwehoeren”, maar daarom ben je wel een hele gezellige en toffe gozer. Joost je bent een goeie, gulle vent en is die Alfa Romeo al gerestaureerd? Eric Milot thank you for your support in the first few years as a PhD student. In het rijtje oud-labgenoten die ik dank verschuldigd ben mag ik Kim (stop toch met “doktertje spelen” en ga het onderzoek in), Rianne (veel succes met je nieuwe opleiding), Mariken, Toleiv, en Stefania niet vergeten. Bedankt voor jullie gezelligheid en steun.

Dear people of 706 it was a great pleasure to have you as neighbours. I enjoyed the sushi and cakes you served, your sharing of knowledge, and our diner and drinks occassions. Thank you John, Elena, Amy, Patrick, Ali, Albert en natuurlijk Edgar.

Jeroen C ik zal bij deze plechtig beloven wat vaker naar de kapper te gaan en de enzymen zijn vanaf nu in goede handen bij Erik. Frustraties over ontbrekende knippers behoort nu dus definitief tot het verleden.

Marieke en Jasperina bedankt voor alle logistieke ondersteuning. Melle en Mieke dankzij jullie is het bestellen van enzymen een aangename taak. Veel succes met Erik.

Buiten het lab waren het Carlo en Maarten die voor de nodige ontspanning zorgden. Hoewel het er de laatste 2 jaar nauwelijks van gekomen is vind ik het fantastisch om samen met jullie koffie te drinken, muziek te maken en bier te drinken (in willekeurige volgorde). Vanaf september ga ik mijn leven beteren en ervoor zorgen dat we weer wat tijd samen door kunnen brengen. Maarten groeten aan Desiree, en Carlo groeten aan Kaatje.

Lieve mam en pap, jullie steun en liefde door alle jaren heen heeft er voor een heel groot deel aan bijgedragen dat nu dit proefschrift hier ligt. Schoonpa en ma jullie interesse en steun de afgelopen jaren was hartverwarmend. Stefan, Tineke, Timo, Thea, Tjeerd, Gerriene, Famke en de rest van de familie bedankt voor jullie steun, interesse, gezelligheid, etc.

Marjon je bent geweldig. Je onvoorwaardelijke liefde betekent oneindig veel voor me. Nu dit proefschrift klaar is gaan we opzoek naar een “nieuw stekkie”, waar we samen met die twee kleine aapies kunnen wonen en leven. Joshua en Anna jullie zijn nog te klein om te begrijpen wat een proefschrift inhoudt, maar pappa heeft jullie gemist de afgelopen maanden. Strakjes gaan we weer samen leuke dingen doen. Marjon, Joshua en Anna ik hou zielsveel van jullie.

Bedankt voor alles,

Bas.