

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

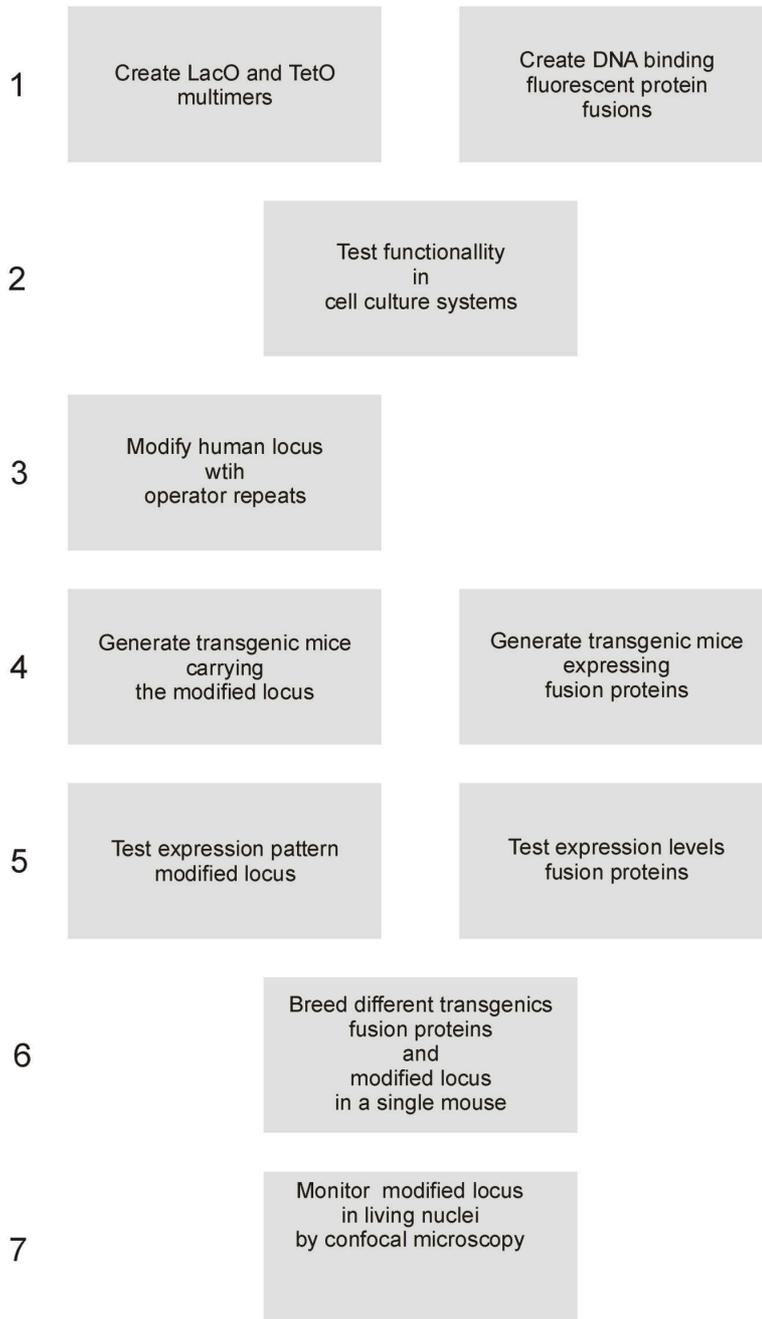
The development of a Lac^O/LacR tagging tool uniquely has allowed examining the dynamics of the chromatin template in living cells [135]. This technique was based on the insertion of 256 direct repeats of the bacterial Lac^O sequence into the genome of cells. Subsequently, cells having undergone gene 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 ($\sim 0.5 \mu\text{m}$ 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 ($\sim 3 \mu\text{m}$ radius) [138]. Thirdly, the morphological structure of large chromatin regions (~ 80 Mbp Lac^O arrays) showed either a fibrillar or a solid density LacR-GFP staining. Electron microscopy showed that the fibrillar 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

Globins in Space

A



B



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 , γ^G , γ^A , 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 $\gamma\delta\beta$ -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 -globin 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).

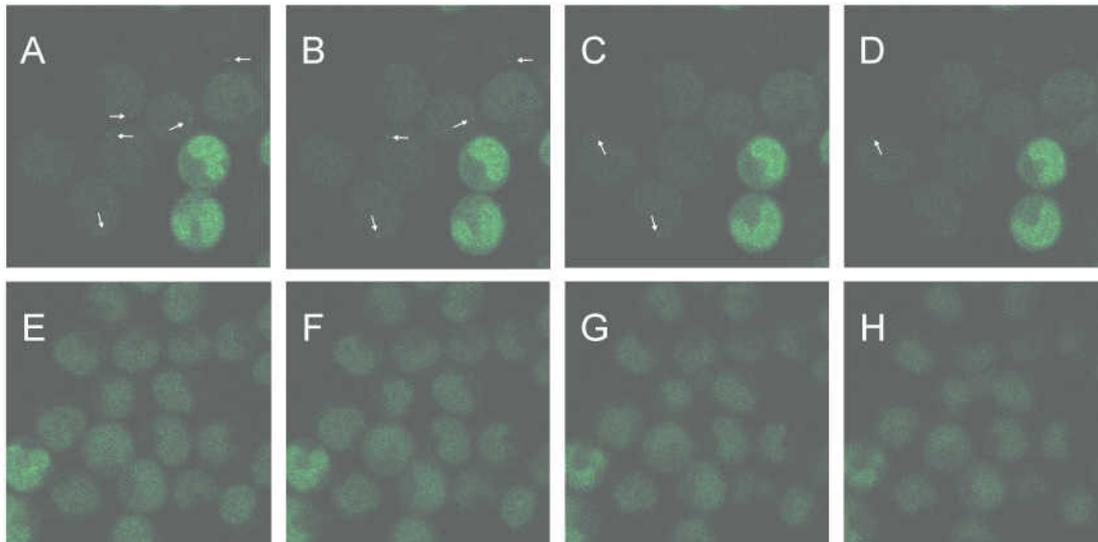


Figure 2.2. A 28-mer of the Tet^O binding sequence can be detected in living cells by confocal microscopy.

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 $\sim 3 \times 10^6$ 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 constructed 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

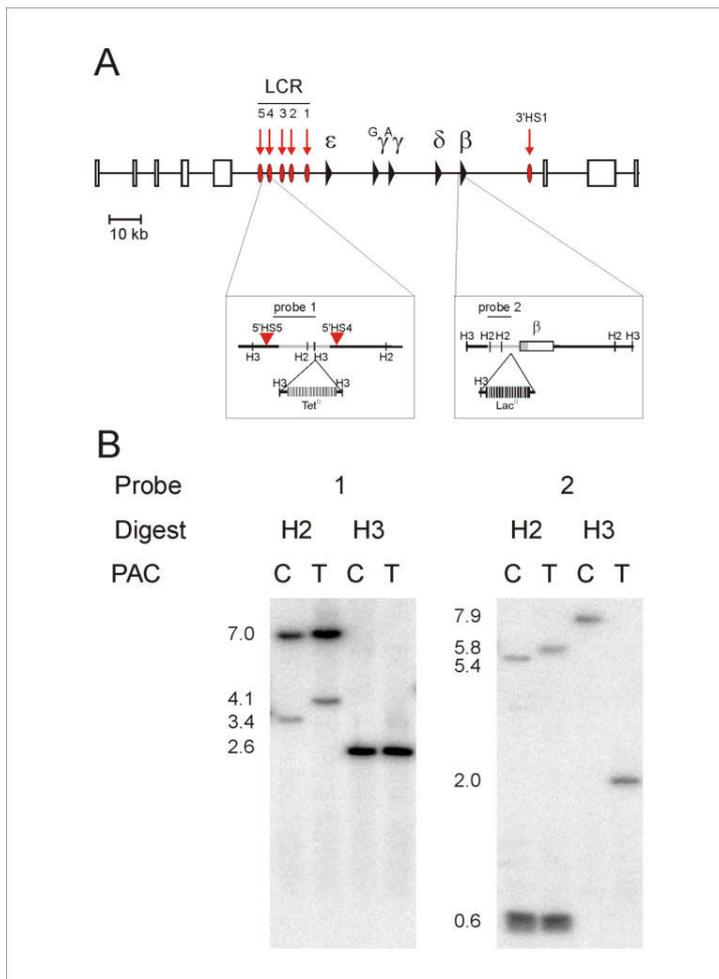


Figure 2.3. Tet^O and Lac^O sequences in PAC TL-O.

A schematic presentation of the PAC148 γ -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: *HindII*; H3: *HindIII*). Southern blot analysis with probe 1 (B left panel, see Methods) gave as expected a shift in a *HindII* digest compared to a control PAC. The *HindIII* digest did not show a shift, because the probe cannot detect the 14xTet^O *HindIII* fragment. The same analysis with probe 2 (B right panel, see Methods) gave an unexpectedly low size shift with a *HindII* digest (see text). However, the Lac^O repeat was inserted, since an additional *HindIII* site reduced the detected fragment size. C and T point out control PAC and PAC TL-O, respectively. Sizes are indicated (kb).

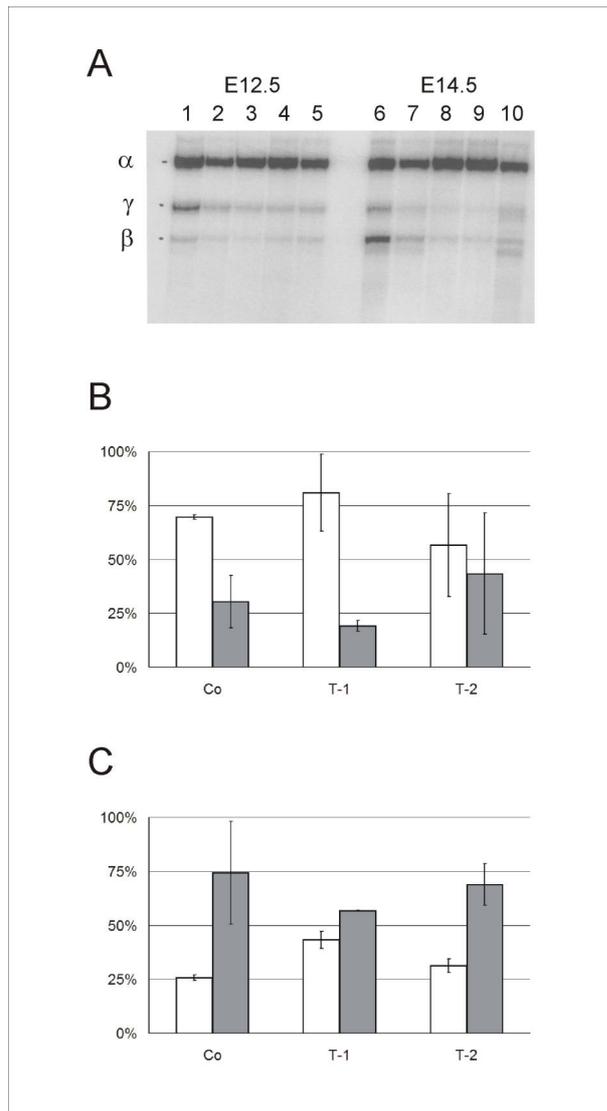


Figure 2.4. Transcription levels in PAC TL-O transgenic mice.

An example of an S1 Nuclease protection assay (A) of $\sim 1 \mu\text{g}$ 12.5dpc fetal liver RNA (lanes 1-5) and $\sim 1 \mu\text{g}$ 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 increased 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 $(\gamma \text{ or } \beta)/(\gamma \text{ and } \beta)$. 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 $\sim 3.2\text{kb}$ *EcoRV/StuI* fragment of the human β -globin LCR with a 14-mer Tet^O repeat ($\sim 0.6 \text{ kb}$) inserted into a *HindIII* 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 *HindIII* integration site, gave a diagnostic shift in a *HindIII* digest compared to a control PAC (figure 2.3B). Subsequently, we used a $\sim 1.9 \text{ kb}$ *BamHI* fragment with a 32-mer Lac^O ($\sim 1.6 \text{ kb}$) sequence inserted into a *BsmBI* site, placing the insert $\sim 400 \text{ bp}$ upstream of the RNA start site (figure 2.3A). However, a *HindIII* 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 \pm 0.13	0.22 \pm 0.02	0.80 \pm 0.21
14.5dpc	1.00 \pm 0.19	0.08 \pm 0.00	0.64 \pm 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 \sim 3 independent S1 Nuclease protection assays and standard-errors are indicated.

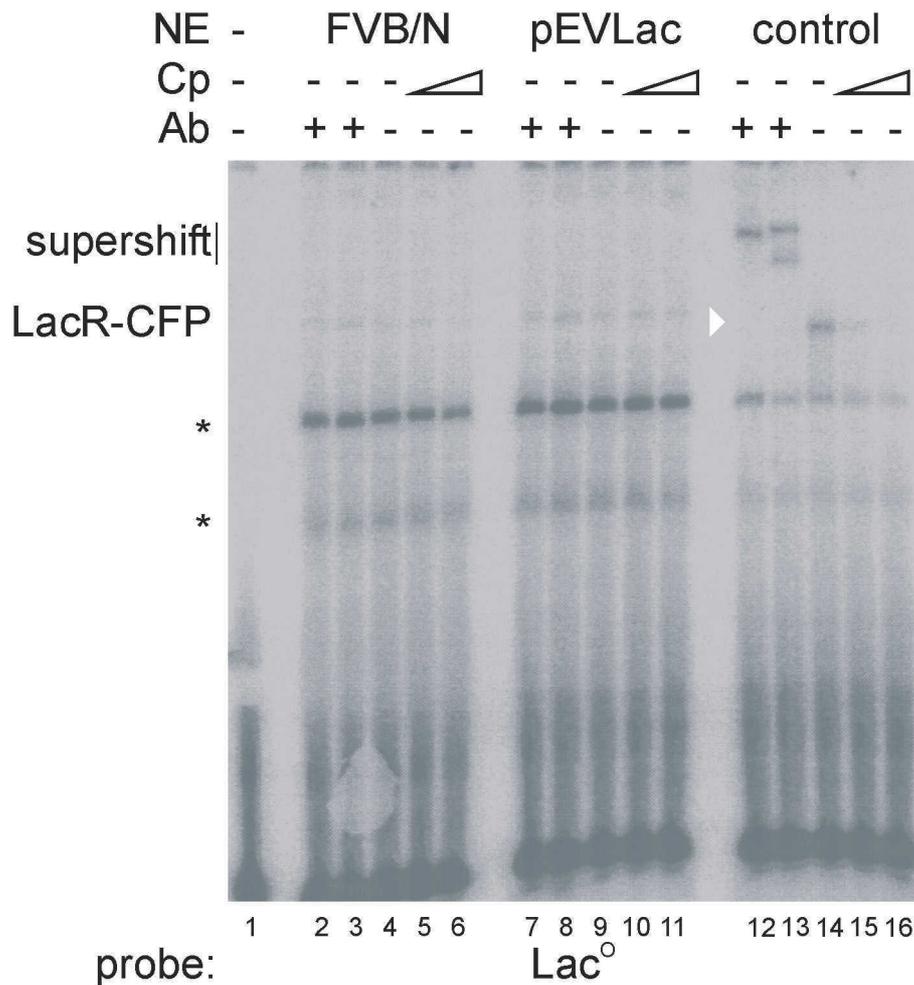


Figure 2.5. Gel mobility shift assay.

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 transgenics are an exception or whether 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 7xTet⁰-construct involved insertion of a 310 bp heptamerized tet operator fragment (*EcoRI*-blunt/*KpnI*) and a pgkPuro (*HindIII*-blunt/*KpnI*) fragment into pBlueScript KS- (Stratagene). The 7xTet⁰ fragment was derived from plasmid pUHC13-3 [243]. The TetR-GFP fusion construct required insertion of SV40NLS (*EcoRI/XbaI*), 618 bp *tet* repressor PCR fragment (*XbaI/SacII*) and GFP (*SacII/NotI*) 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 (*XbaI/SacII*) 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 14xTet⁰ vector contained a ~3.2kb *EcoRV/StuI* fragment of the human β -globin LCR with two tandem 7xTet⁰ repeats (~0.6 kb) inserted into its *HindIII* site. This ~3.8 kb fragment was inserted *EcoRV/StuI* blunt into a *HindIII* site of pDF25 [232]. For the generation of the 32xLac⁰ recombination vector we first amplified a single Lac⁰ to a 32-mer. We used an insert amplification strategy analogous to Robinett et al. [135]. Briefly, a sense and antisense oligomer containing a single Lac⁰ were annealed and inserted into a *SalI* site of pBlueScript SK+ (Stratagene). A subamount of this pLac⁰1x, including the operator, was digested with *SacII/ClaI* and another amount was digested with *SacII/AclI*, 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 *ClaI* and *AclI*, which after ligation could not be recut with either enzyme. Six rounds of this cloning cycle generated a pLac⁰32x plasmid. The 32xLac⁰ recombination vector (pDF25) consisted of a ~1.9 kb *BamHI* β -globin promoter fragment with the 32x Lac⁰ sequence inserted *HinIII/EcoRV* into a blunt *BsmBI* site. The homologous recombination of PAC148 γ lox has been described before [232]. Lac⁰ 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 *SstI/HindIII* fragment surrounding 5' HS5 of the human β -globin locus; probe 2 is a ~1.0kb *BamHI/BsmBI* 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 μ l 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 continuous CO₂ 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 1 μ m were taken using an Axiovert inverted microscope with Axiovert lens 63 x oil (1.4 n.a., working distance 90 μ m), 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 *NotI* 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.

AatII/Asp718I digestion isolated the 11.2 kb TetR-YFP and 11.6 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/ μ l 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 β IVS2 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 LCR ϵ and $\gamma\gamma\delta\beta$ cosmid probes and specific probes within the LCR and the different genes. Transgene copy numbers were determined using as probes the β IVS2 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 β IVS2 /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 *Avall* 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 *BamHI* fragment, protected fragment size 185 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 μ g of nuclear extracts for 20 minutes at room temperature in 10 mM Tris pH7.5; 10 mM MgCl₂; 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.