# 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].

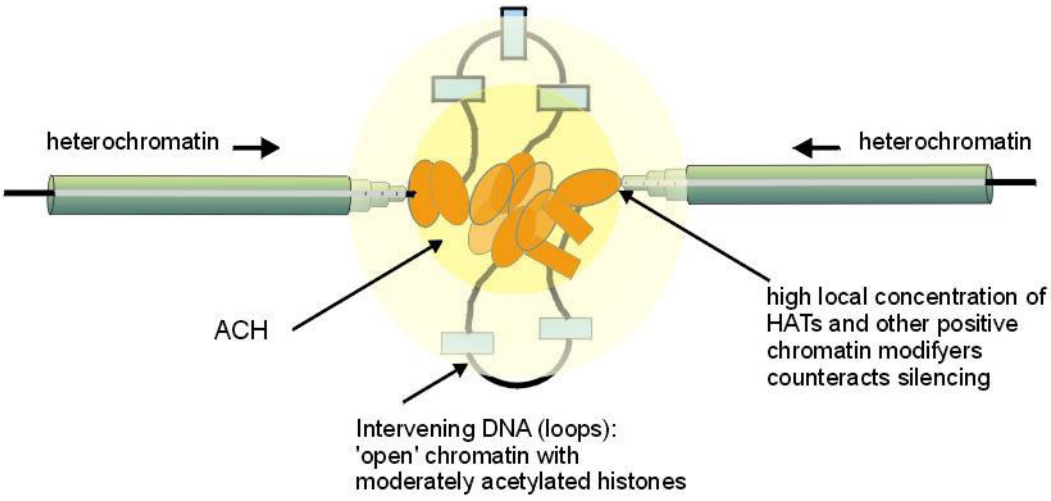


**Figure 5.1. Hypothetical: the ACH and transcription in repressive chromatin.**

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

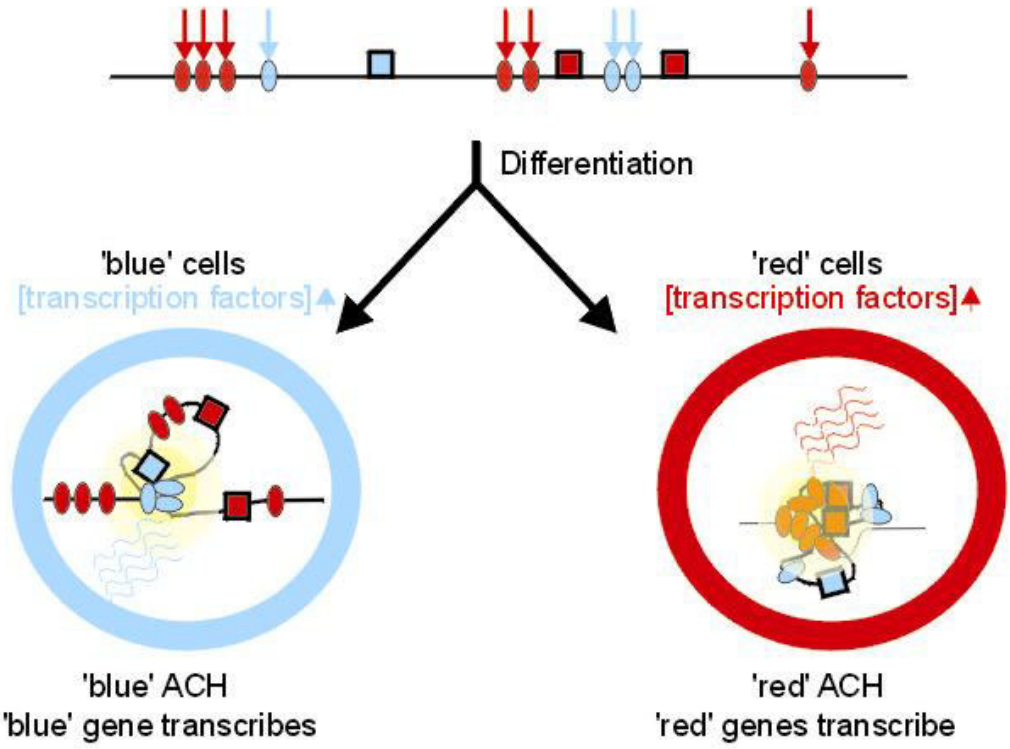


**Figure 5.2. Hypothetical: the ACH and expression of overlapping gene loci.**

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.