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

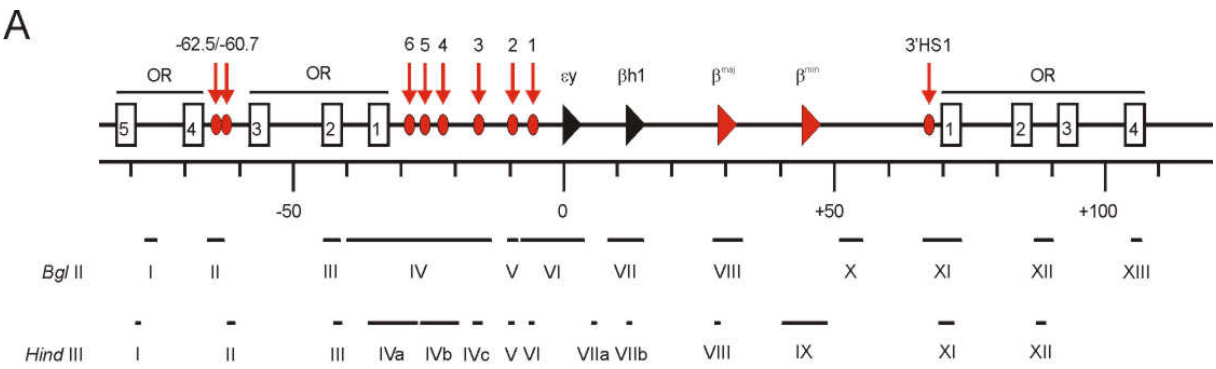
Applying 3C technology to the murine β-globin locus

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*.

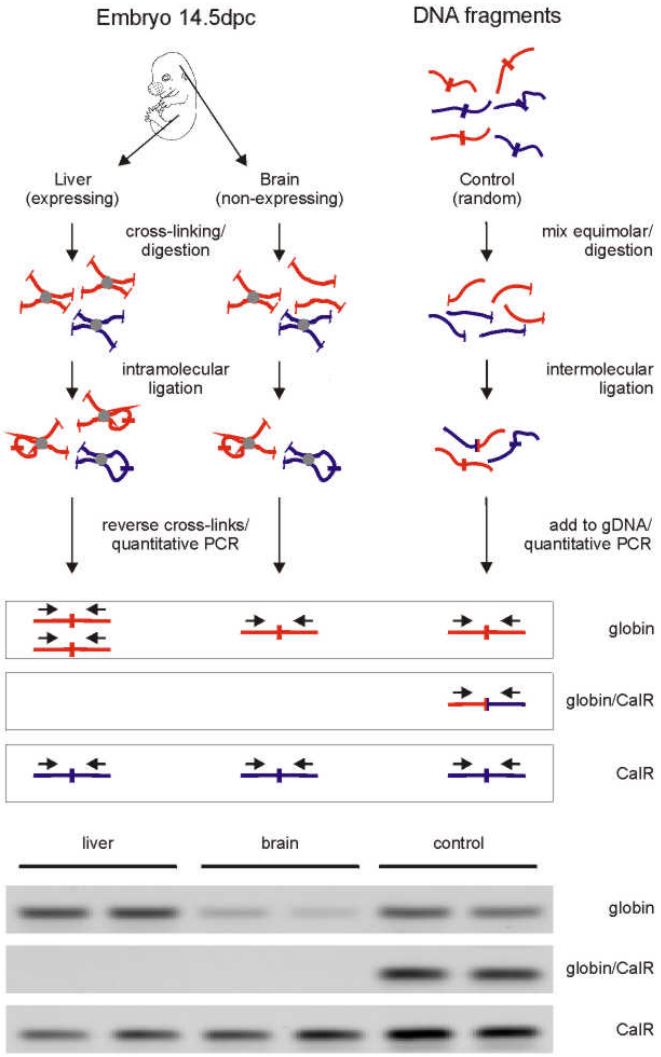
Figure 3.1. 3C technology in the murine β-globin locus

(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.

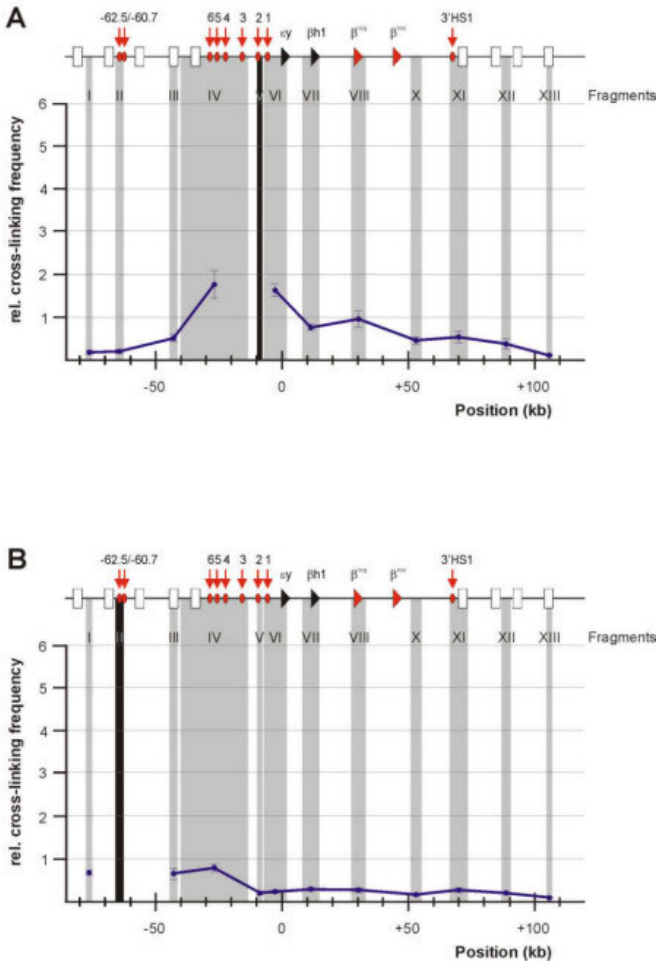


Figure 3.2. Linear conformation of the β-globin locus in non-expressing brain cells.

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.

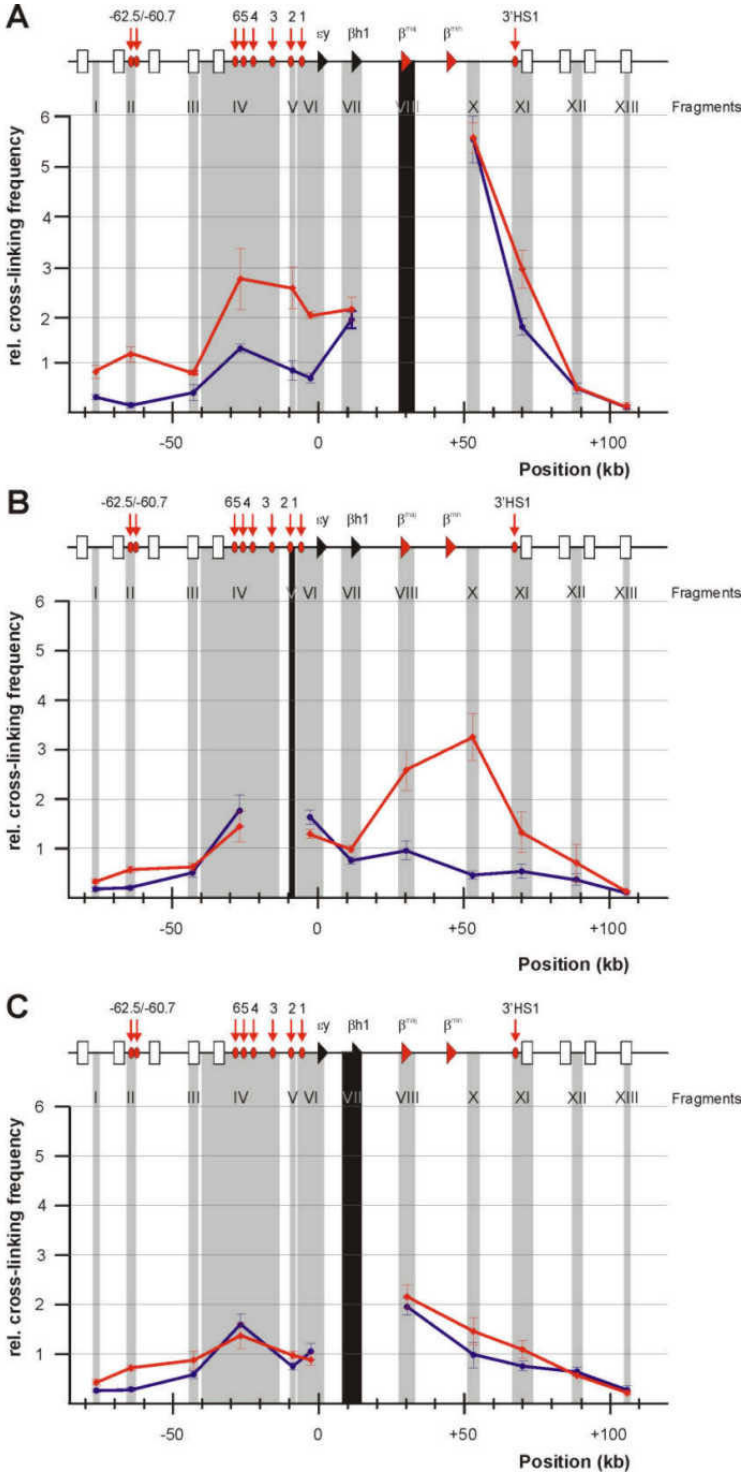


Figure 3.3. Erythroid-specific interaction and looping between the LCR and an active β-globin gene.

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.

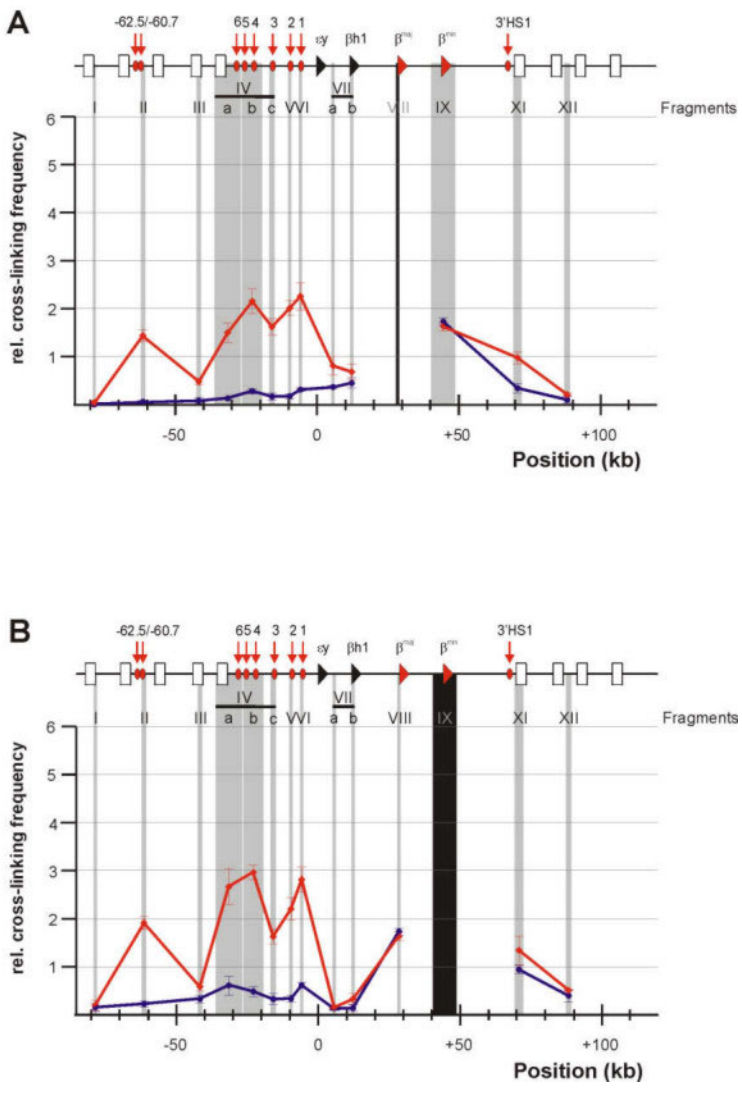


Figure 3.4. Erythroid-specific interactions between the active β-globin genes and individual hypersensitive sites in the LCR.

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-

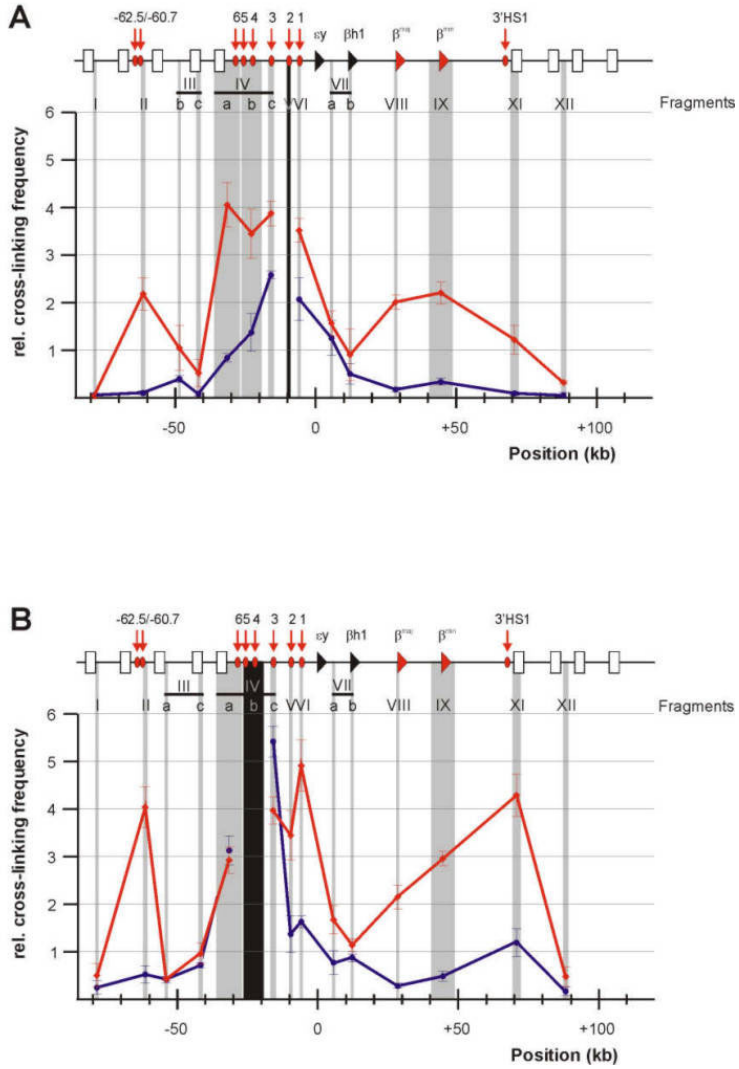


Figure 3.5. Erythroid-specific high cross-linking frequencies among the individual hypersensitive sites of the LCR and two distal hypersensitive sites.

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].

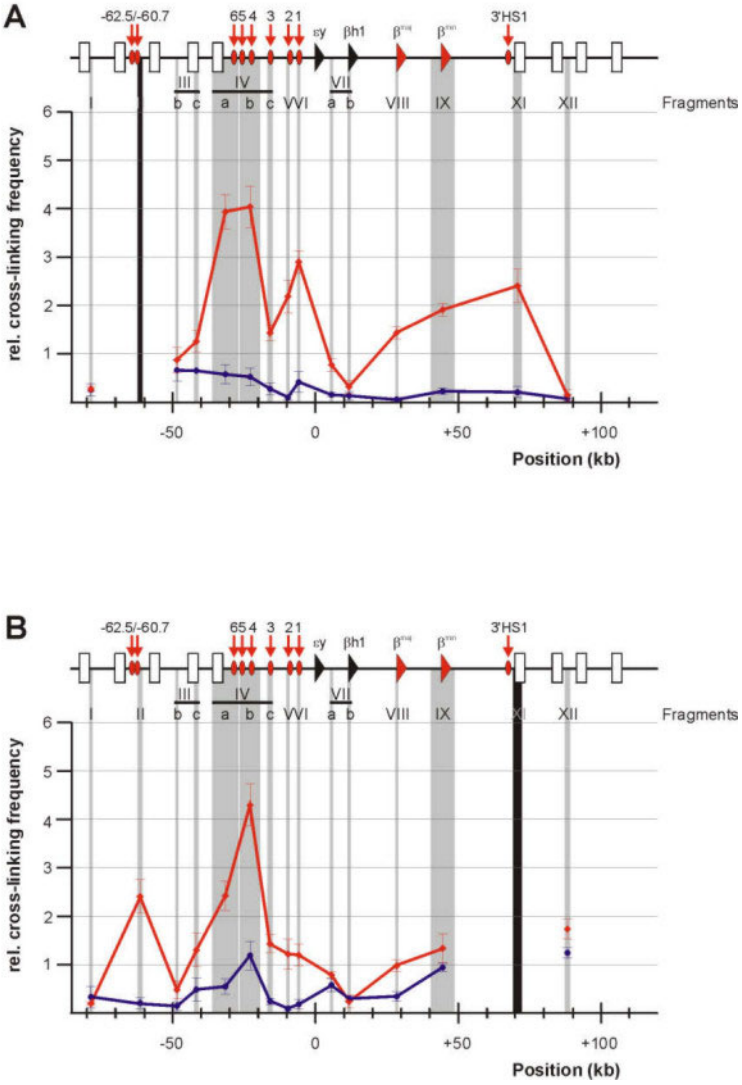


Figure 3.6. Two distal hypersensitive sites at each side of the locus cluster with the LCR and the genes.

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.

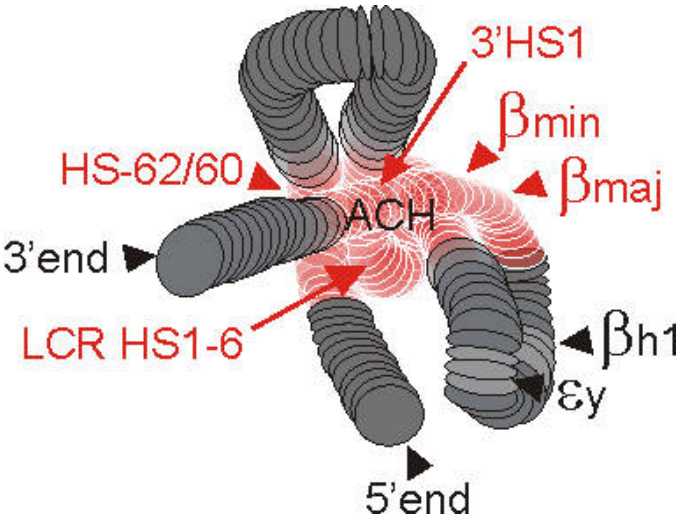


Figure 3.7. A 3D model of the ACH.

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.