DNA-binding factor CTCF and long-range gene interactions in V(D)J recombination and oncogene activation

Claudia Ribeiro de Almeida, *Ralph Stadhouders, Supat Thongjuea, Eric Soler, and Rudi W. Hendriks

Departments of 1Pulmonary Medicine and 2Cell Biology and Genetics, Erasmus Medical Center (MC) Rotterdam, Rotterdam, The Netherlands; 3Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, Bergen, Norway; and 4Cancer Genomics Center, Erasmus MC Rotterdam, Rotterdam, The Netherlands

Regulation of V(D)J recombination events at immunoglobulin (Ig) and T-cell receptor loci in lymphoid cells is complex and achieved via changes in substrate accessibility. Various studies over the last year have identified the DNA-binding zinc-finger protein CCCTC-binding factor (CTCF) as a crucial regulator of long-range chromatin interactions. CTCF often controls specific interactions by preventing inappropriate communication between neighboring regulatory elements or independent chromatin domains. Although recent gene targeting experiments demonstrated that the presence of the CTCF protein is not required for the process of V(D)J recombination per se, CTCF turned out to be essential to control order, lineage specificity and to balance the Ig V gene repertoire. Moreover, CTCF was shown to restrict activity of C enhancer elements to the Ig C locus. In this review, we discuss CTCF function in the regulation of V(D)J recombination on the basis of established knowledge on CTCF-mediated chromatin loop domains in various other loci, including the imprinted H19-Igf2 locus as well as the complex β-globin, MHC class II and IFN-γ loci. Moreover, we discuss that loss of CTCF-mediated restriction of enhancer activity may well contribute to oncogenic activation, when in chromosomal translocations Ig enhancer elements and oncogenes appear in a novel genomic context.

Introduction

Development of B and T lymphocytes requires proper spatial and temporal control of gene expression. This is achieved through the coordinate binding of lineage-specific transcription factors to cis-regulatory elements. In many developmentally regulated gene clusters (eg, the antigen receptor and cytokine loci), gene regulation is particularly complex and involves cell type-specific interactions over large genomic distances between gene promoters and remote enhancers, silencers, or locus control regions (LCRs).

Regulation of higher-order chromatin structures is mediated by protein complexes composed of lineage-specific and ubiquitously expressed nuclear factors. It has recently become apparent that the DNA-binding protein CCCTC-binding factor (CTCF) is a leading candidate for the regulation of gene expression at complex loci during lymphoid differentiation.

General introduction into CTCF biology

CTCF is a ubiquitously expressed and highly conserved 11-zinc finger protein that was originally identified as a negative transcriptional regulator of the c-Myc oncogene. CTCF has since been implicated in a variety of regulatory functions, including transcriptional activation and repression, enhancer blocking, gene insula-
tion, gene silencing, genomic imprinting, X chromosome inactivation, and long-range chromatin interactions.

CTCF occupancy studies in various human and murine cell types, including embryonic stem cells and T lymphocytes, revealed approximately 14 000 to approximately 40 000 CTCF-binding sites, sharing a 15- to 20-bp core consensus sequence. Although the genome-wide distribution of these sites strongly correlates with gene density, it suggests a role for CTCF distinct from that of canonical transcription factors: binding patterns do not correlate with transcriptional start sites and do not appear to predict cell-specific gene expression levels. Domains with few or no CTCF-binding sites tend to include clusters of transcriptionally coregulated genes that are often flanked by CTCF-binding sites. Conversely, CTCF-rich regions contain genes displaying extensive alternative promoter usage, including the T-cell receptor (TCR) and immunoglobulin (Ig) gene loci.

Many studies have suggested that CTCF is the archetypal vertebrate protein that binds insulator elements. Insulators affect gene expression by blocking inappropriate communication between neighboring regulatory elements (enhancer-blocking activity, when positioned between enhancer and promoter) and by preventing the spread of repressive heterochromatin (barrier function). Based on studies showing that a 5′ boundary element of the chicken β-globin locus serves as an insulator in erythroid cells and protects against position effect in Drosophila, Chung et al were the first to hypothesize that insulator function may depend on interactions between distant elements that result in looping out of intervening sequences. The first association of CTCF with enhancer blocking activity was proposed on the basis of CTCF-binding to the DNAse hypersensitive site (5′HS4) upstream of the chicken β-globin locus and another insulator sequence at the 3′ end. Based on functional analysis in transgenic systems, CTCF has been demonstrated to mediate insulation at several model loci discussed below in detail. In addition, computational analyses showing CTCF binding to highly conserved noncoding elements

© 2012 by The American Society of Hematology


* C.R.d.A. and R.S. contributed equally to this study.
across 12 mammalian species provided evidence supporting a global role for CTCF as an insulator protein.12

CTCF-mediated gene insulation via chromatin loops in model loci

CTCF drives enhancer-blocking insulation through long-range chromatin interactions, in particular through 3-dimensional loop formation between distal elements.4,13

The H19/Igf2 locus

CTCF has been implicated in genomic imprinting, the epigenetically regulated process that causes genes to be expressed in a parental-origin–specific manner rather than from both chromosomes.14 In the H19/Igf2 locus, an imprinting control region, containing 4 CTCF-binding sites, is located immediately upstream of the H19 gene and is essential for regulation of imprinted expression of the H19 and Igf2 genes (Figure 1A). On the maternally inherited allele, the imprinting control region is unmethylated, allowing for CTCF binding and the formation of a tightly coiled chromatin loop insulating the Igf2 gene and preventing its promoter to access the enhancers downstream of H19. On the paternal allele, however, the imprinting control region is methylated, which abrogates CTCF binding, resulting in functional interactions between the Igf2 promoters and the enhancer elements (Figure 1A). Moreover, loss of CTCF binding is also associated with H19 promoter methylation leading to H19 gene silencing on the paternal allele.

The β-globin locus

CTCF-mediated contacts are associated with transcriptional activation of genes in the extensively characterized β-globin locus in erythroid cells.15 Developmentally regulated expression of individual globin genes is controlled by many cis-acting elements, the most important of which are located in the upstream LCR (Figure 1B). CTCF binds to 3 DNAse hypersensitive sites upstream in the LCR and 1 downstream (3’HS1) of the mouse β-globin locus in a
cell type–specific manner. During erythroid differentiation, CTCF-bound regulatory sequences throughout the locus come into close spatial proximity to form an “active chromatin hub,” as was shown by chromosome conformation capture (3C) technology detecting long-range DNA interactions. Formation of this “active chromatin hub” allows stage-specific interaction of active globin genes with the LCR and concomitant looping-out of transcriptionally silent isoforms (Figure 1B). Conditional CTCF deletion or mutation of the 3′HS1 CTCF-binding site resulted in disruption of CTCF contacts in erythroid progenitors but surprisingly had no effects on kinetics or levels of globin expression during erythroid differentiation. Because of the high density of CTCF-binding sites in the β-globin locus, this finding is explained by redundancy of chromatin contacts, which may not be abrogated by mutations in a single regulatory element. Genetic disruption of CTCF-binding to 3′HS1 also did not affect transcription of the neighboring olfactory receptor genes. Collectively, these findings indicate that correct globin gene expression can be achieved by different CTCF-dependent loops and thus by different chromatin conformations.

The MHC class II locus

CTCF organizes the MHC class II locus into a conformation that rearranges in the presence of the crucial non-DNA–binding transcriptional coactivator CIITA. In the human locus, CTCF binds to the intergenic enhancer XL9 that coregulates expression of the HLA-DRB1 and HLA-DQA1 genes (Figure 1C). Induction of CIITA in non-MHC class II–expressing cells results in the initiation of interactions between XL9 and the promoters upstream of HLA-DRB1 and HLA-DQA1. Hereby chromatin loop formation depends on XL9-bound CTCF, CIITA, as well as the transcription factor RXF, which is constitutively bound to regulatory sequences within the HLA-DRB1 and HLA-DQA1 proximal promoters. Therefore, in the human MHC class II locus, CTCF is involved in the control of gene expression through loop formation via hetero-multimerization.

The Ifng and Th2 cytokine loci

The key transcription factor of T helper-1 (Th1) cells, T-bet, promotes the expression of IFN-γ in part by facilitating CTCF binding in the Ifng locus. CTCF acts to establish a Th1 cell–specific Ifng locus architecture, which helps to drive the juxtaposition of T-bet–binding enhancers, the flanking CTCF-binding elements, and CTCF sites in the first exon of the Ifng gene (Figure 1D). Cooperation between CTCF and T-bet is required for robust Th1 cell–specific IFN-γ expression, which is also supported by our finding of an approximately 50% reduction in IFN-γ–producing cells when naïve CTCF-deficient CD4+ T cells were cultured under Th1-polarized conditions. CTCF may act in a similar fashion in the Th2 cytokine locus containing the genes for IL-4, IL-5, and IL-13. Indeed, Th2 cytokines were strongly reduced in CTCF-deficient T lymphocytes cultured under Th2-polarized conditions, despite proper induction of the key Th2 transcriptional regulators Gata3 and Satb1.

CTCF as a genome-wide organizer of chromatin architecture

Its capacity to form dimers and to participate in multiprotein complexes allows CTCF to organize long-range 3-dimensional looping of genomic regions. Interestingly, genome-wide identifica-
Igκ/Igλ chain rearrangement results in surface expression of the BCR and additional checkpoints follow to limit autoreactivity.32 Likewise, in the thymus, V(D)J recombination events in the TCR loci accompany the development of mature T cells, expressing an αβ TCR or, in a minor subset, a γδ TCR on the cell surface.32

Given that all Ig and TCR loci are rearranged by a common recombination machinery, an essential level of regulation of the V(D)J recombination process relies on lineage and developmental stage-specific changes in the accessibility of antigen receptor loci for recombination. The “accessibility hypothesis” was proposed based on the initial observation that rearranging gene loci for recombination. The “accessibility hypothesis” was proposed based on the initial observation that rearranging gene segments is concomitantly undergoing germline transcription.28,33 It has been demonstrated that transcription itself can directly recombine targeting: blockade of transcriptional elongation through the mouse TCR-α locus suppressed Vα to Jα recombination and chromatin remodeling of Jα segments.34 Germline transcription increases recombination of Jα segments located close to an upstream crucial promoter, called TEA, and prevents activation of downstream promoters through transcriptional interference.34 Accessibility is controlled by local cis-regulatory elements (promoters and enhancers) and involves sub-nuclear relocation, histone acetylation and/or methylation, DNA demethylation, and locus compaction.1,28 These processes are dependent on various lineage-specific transcription factors that bind to enhancer elements present in the Ig and TCR loci.35-37 Among these transcription factors, Pax5, EBF, and E2A play a central role in B-cell differentiation, as they also establish a hierarchical regulatory network for specification and commitment to the B-cell fate.38 Next to CTCF, only a few general nuclear factors that modulate antigen receptor loci topology have been identified to date, including YY1 and Ezh2.35,37,39,40 However, the precise mechanism by which these factors act in concert with lineage-specific transcription factors to regulate chromatin structure at the Ig loci remains unclear.

CTCF function in IgH locus gene rearrangement

The IgH locus in the mouse spans an approximately 2.7-Mbp region and is composed of more than 100 VH gene segments and Dδ, Jδ, and Cδ clusters located downstream. The locus contains 2 important regulatory elements: the intronic enhancer (iEδ), located 3' of the JH genes, and the 3' regulator region (3'RR; Figure 2A).28 The stepwise activation of the locus, reflected by V-ho-JH recombination, is paralleled by germline transcription. First, iEδ transcripts are generated from iEδ, μ0 transcripts from the promoter of the most 3' Dδ segments, and antisense intergenic transcripts throughout the Dδ and JH domains.28,41,42 After Dδ-JH recombination, noncoding sense and antisense transcripts are generated from the VH genes.28,33,42,43 Three-dimensional FISH experiments have elucidated why pro-B cells deficient for Pax5, YY1, or Ikaros only recombine D-proximal VH genes, suggesting that locus compaction mediated by these proteins is necessary to bring distal VH genes in close proximity to Dδ-JH.35,37,39 Spatial distance measurements of genomic markers spanning the entire IgH locus have demonstrated that this locus is organized into compartments that contain clusters of loops or rosette-type structures separated by linkers.44 This topology compacts the VH region and should ensure that VH gene segments scattered over approximately 2.5 Mbp encounter Dδ-JH segments or enhancers with appropriate frequencies.

Emerging data have revealed CTCF as an attractive candidate for controlling IgH locus topology. CTCF and cohesin bind to multiple sites throughout the VH region,45 implicating CTCF in IgH...
chromatin loop formation and effective distal V\textsubscript{H}-D\textsubscript{H}J\textsubscript{H} recombination.\textsuperscript{46,47} However, CTCF occupancy in the V\textsubscript{H} region is largely unchanged between pro-B and pre-B cells,\textsuperscript{45} making it unlikely that CTCF alone accounts for Ig\textsubscript{H} locus topology changes. Consistently, 3C-based experiments demonstrated that shRNA-mediated CTCF deletion only modestly affected Ig\textsubscript{H} locus contraction.\textsuperscript{48} In agreement with these findings, on mb1-Cre–mediated conditional deletion of the Ctcf gene in the B-cell lineage, Ig\textsubscript{M}+ pre-B cells were still generated and both distal and proximal V\textsubscript{H} gene segments were used.\textsuperscript{49}

Using chromatin immunoprecipitation coupled to microarray analysis (ChIP-chip) technology, Ebert et al have mapped Ig\textsubscript{H} active histone modifications and transcription factor occupancy.\textsuperscript{50} They hereby identified 14 copies of a potential regulatory Pax5-activated intergenic repeat (PAIR) element in the distal V\textsubscript{H} region, upstream of V\textsubscript{H}3609 genes and interspersed among V\textsubscript{H}558 genes. These PAIR elements contain Pax5-dependent active chromatin and binding sites for E2A, CTCF, and the cohesin component Rad21. Detailed analysis of 3 PAIR elements showed that immediately upstream of the CTCF-binding sites noncoding antisense transcription is initiated exclusively in pro-B cells. The authors hypothesize that PAIR elements mediate looping of distal V\textsubscript{H} genes in pro-B cells, whereby Pax5 orchestrates the formation of a bridging complex that mediates long-range interactions between distal and proximal V\textsubscript{H} gene regions. In contrast, on the basis of anti-CTCF Chip-loop assays, Guo et al\textsuperscript{51} presented evidence for a CTCF-mediated but iE\textsubscript{\alpha}-independent multiloop domain of approximately 500 kb in the 5' V\textsubscript{H}558/V\textsubscript{H}3609 region containing the PAIR elements, whereby locus contraction may result from interaction between E\textsubscript{\mu}-bound\textsuperscript{49} Yy1 and Yy1 sites in the distal V\textsubscript{H} region. These interactions may be increased by heterotypic interactions with CTCF sites or by inter-PAIR interactions in the distal V\textsubscript{H} region.\textsuperscript{51} Future experiments should focus on functional characterization of the PAIR elements, which would require deletion of these proposed regulatory elements to test the effects on V(D)J recombination and V\textsubscript{H} repertoire in vivo.

In addition, CTCF-binding sites are present in the intergenic control region I (IGCR1), which is located 3 to 5 kb upstream of the most 5’ D\textsubscript{H} element (Figure 2A).\textsuperscript{45,52,53} These CTCF-binding sites have enhancer-blocking activity and mark a sharp decrease in E\textsubscript{\mu}-dependent antisense D\textsubscript{H}J\textsubscript{H} transcription.\textsuperscript{52} Targeted IGCR1 in the mouse leads to D\textsubscript{H}J\textsubscript{H} antisense transcripts extending into proximal V\textsubscript{H} segments and concomitant aberrant rearrangement of the proximal V\textsubscript{H} gene segments in developing thymocytes.\textsuperscript{53,54} Importantly, IGCR1 balances proximal versus distal V\textsubscript{H} usage by inhibiting germline transcription and rearrangement of D-proximal V\textsubscript{H}7183 gene segments (Figure 2A). Moreover, IGCR1 maintains ordered V(D)J recombination by suppressing V\textsubscript{H} joining to D segments that are not joined to J\textsubscript{H}. Both FISH and 3C-based experiments showed that CTCF-binding sites in IGCR1 are in close spatial proximity with a region immediately downstream of 3’RR with a high density of CTCF-binding sites (Figure 2A).\textsuperscript{48,51,55} These CTCF-mediated interactions would create a domain that contains all of the D\textsubscript{H}, J\textsubscript{H}, and C region elements, as well as E\textsubscript{\mu} and 3’RR. It therefore seems probable that CTCF binding to IGCR1 prevents chromatin accessibility extending into the independently activated V\textsubscript{H} chromatin domains. This notion is supported by the findings that (1) insertion of a V\textsubscript{H} gene segment just 5’ of the D\textsubscript{H} cluster resulted in loss of ordered and lineage-specific Ig\textsubscript{H} locus rearrangement,\textsuperscript{56} and (2) mice with an approximately 8-kb deletion encompassing the CTCF-binding sites downstream 3’RR showed a 2-fold increase in the usage of V\textsubscript{H}7183.\textsuperscript{57}

In summary, by binding to IGCR1 and downstream of 3’RR, CTCF acts as an insulator that ensures ordered and lineage-specific V(D)J recombination at the Ig\textsubscript{H} locus.

**CTCF function in Ig\textsubscript{\kappa} locus gene rearrangement**

The Ig\textsubscript{\kappa} locus in the mouse spans approximately 3.2 Mbp and contains approximately 100 V\textsubscript{\kappa} gene segments, a J\textsubscript{\kappa} cluster and a single C\textsubscript{\kappa} region.\textsuperscript{20} Ig\textsubscript{\kappa} locus rearrangement and expression are positively regulated by an intronic enhancer (iE\textsubscript{\kappa}) located between J\textsubscript{\kappa} and C\textsubscript{\kappa}, and 2 enhancers downstream of C\textsubscript{\kappa} (3’E\textsubscript{\kappa} and Ed; Figure 2B). As D elements are lacking, \kappa light chains can be expressed on a single productive V\textsubscript{\kappa}-to-J\textsubscript{\kappa} recombination event and autoactive BCR specificities can be replaced by “receptor editing,” the process of ongoing Ig light chain gene recombination.\textsuperscript{128} Germline \kappa\textsuperscript{0} transcripts from promoters located upstream of J\textsubscript{\kappa} (\kappa\textsuperscript{0} 1.1 and \kappa\textsuperscript{0} 0.8) or V\textsubscript{\kappa} gene segments indicate the accessibility over J\textsubscript{\kappa} and V\textsubscript{\kappa} regions, respectively.\textsuperscript{58} A recombination silencer element in the V\textsubscript{\kappa}-J\textsubscript{\kappa} region, called “silencer in intervening sequence” (SIS), negatively regulates Ig\textsubscript{\kappa} rearrangement by targeting of the nonrearranging allele to centromeric heterochromatin (Figure 2B).\textsuperscript{59} Deletion of the SIS element resulted in decreased monoallelic Ig\textsubscript{\kappa} heterochromatin localization, reduced \kappa\textsuperscript{0} germline transcription, and enhanced proximal V\textsubscript{\kappa} usage.\textsuperscript{59,60}

Next to strong CTCF binding at the 5’ and 3’ boundaries and the SIS element,\textsuperscript{49} we identified approximately 60 CTCF-binding sites in the V\textsubscript{\kappa} region (in contrast to the previously reported low density of CTCF occupancy in the Ig\textsubscript{\kappa} locus).\textsuperscript{45} CTCF-binding sites were not evenly distributed over the Ig\textsubscript{\kappa} locus. Detailed analysis of long-range interactions in pre-B cells revealed a correlation between V\textsubscript{\kappa} usage, local density of CTCF-binding sites, and the frequencies of contacts with SIS, iE\textsuperscript{\kappa}, or 3’E\textsuperscript{\kappa}.\textsuperscript{61} Using 3C-seq, we observed that long-range interactions between the SIS-, iE\textsuperscript{\kappa}-, or 3’E\textsuperscript{\kappa}- and V\textsubscript{\kappa} region were not uniformly affected by conditional deletion of the CTCF gene in the B-cell lineage. Loss of CTCF resulted in significantly increased interactions with the most proximal V\textsubscript{\kappa} region containing the V\textsubscript{\kappa} family. In line with these findings, CTCF-deficient pre-B cells showed significantly increased V\textsubscript{\kappa} germline transcription and recombination.\textsuperscript{59} On the other hand, loss of CTCF resulted in reduced contacts with the 2 most distal V\textsubscript{\kappa} genes, V\textsubscript{\kappa}2-137 and V\textsubscript{\kappa}1-135, which are often rearranged in control but not in CTCF-deficient (pre-B) cells.

Based on these studies,\textsuperscript{49,59,60} we proposed that CTCF binding in the SIS element provides enhancer-blocking activity that limits interactions between the enhancers and the proximal V\textsubscript{\kappa} promoters. CTCF, however, plays a dual role, as, next to the enhancer-blocking activity, it tethers remote V\textsubscript{\kappa} gene promoters to the enhancers.

**Comparison of CTCF function in the Ig\textsubscript{\kappa} and Ig\textsubscript{\gamma} locus**

The enhancer blocking activity of CTCF at the SIS element in the V\textsubscript{\kappa}-J\textsubscript{\kappa} intergenic region in the Ig\textsubscript{\kappa} locus\textsuperscript{49,60} remarkably parallels the role of CTCF at IGCR1 in the V\textsubscript{\gamma}-D intergenic region in the Ig\textsubscript{\gamma} locus.\textsuperscript{53} Likewise, in thymocytes, CTCF colocalizes with cohesin at the TCR-\alpha regulatory elements TEA just upstream of J\textsubscript{\alpha}. Loss of cohesion in double-positive
thymocytes led to reduced long-range TEA-\(\alpha\) interactions and significantly decreased germline transcription and rearrangement over the 3'\(J_\alpha\) gene segments.\(^{61}\) Cohesin and CTCF also colocalize at 5' boundary sites just downstream of the E\(\alpha\) enhancer, demarcating the TCR-\(\alpha\) locus from the neighboring gene Dad1,\(^{61}\) a region with CTCF-dependent insulator activity.\(^{62}\) On virtually complete deletion of CTCF protein in the B-cell lineage in our conditionally targeted mice, we found that productive Igh and Igk rearrangements were still generated, demonstrating that CTCF is not required for V(D)J recombination per se. This is also consistent with the presence of TCR-\(\alpha\)- or TCR-\(\beta\)-chain rearrangement in CTCF-deficient thymocytes in vivo.\(^{63}\) The Igh and Igk loci both contain approximately 60 CTCF-binding sites that influence the genomic architecture of these loci. The identification in the Igh locus of PAIR elements just upstream of distal V\(H_3\)609 genes,\(^{50}\) together with the observation that many proximal V\(H\) genes segments are located within 100 bp of CTCF-binding sites,\(^{64}\) suggests that proximity of a CTCF-binding site may affect the probabilities of individual V\(H\) genes to encounter a DH-J\(H\) element for recombination. In this aspect, the Igh and Igk loci are remarkably different: whereas approximately 27% of V\(H\) genes have a CTCF-binding site within 1-kb distance, this is the case for only 2 V\(\kappa\) genes (Figure 3). Moreover, only approximately 11% of V\(H\) genes but approximately 36% V\(\kappa\) genes do not have a CTCF-binding site in the intervening sequences up to their 5' or 3' neighboring V genes. Thus, although CTCF binds at many sites throughout the Igk locus, most CTCF occupancy is located between V\(\kappa\) segments and not adjacent to the V\(\kappa\) segments.\(^{1,49}\) It has been suggested that V regions are organized as rosettes by CTCF, whereby CTCF-binding adjacent to a V gene increases its recombination probability.\(^{64}\) Because the Igh and Igk loci show large differences in the proximity of CTCF-binding sites to V genes, in such a model the 2 loci would be very differently organized in 3-dimensional space to provide appropriate access of individual V regions to the proposed recombination center.\(^{65}\) Interestingly, it was recently proposed that E2A proteins may modulate Igk locus topology by acting as anchors in a mechanism similar to that put forward for CTCF in the Igh locus.\(^1\) This hypothesis still needs to be tested but is supported by the remarkable, nonrandom distribution of E2A binding sites across the Igk locus, in particular within 200 bp of the 5' or the 3' end of V\(\kappa\) regions.\(^{1,46,47}\)

CTCF and long-range gene interactions in Ig enhancer-mediated oncogene activation

Many leukemias and lymphomas contain chromosomal translocation involving Ig or TCR loci whereby the strong enhancers in these loci are placed in a new genomic context. Because CTCF and cohesin have been implicated in interchromosomal interactions, they probably contribute to physical proximity of particular translocation targets in lymphocytes. For example, in many Burkitt B-cell lymphomas, translocations place the MYC oncogene up to approximately 200 kb upstream of the Igh 3'\(RR\) (Figure 4A). In normal B cells, Igh and MYC are preferentially positioned in close spatial proximity relative to each other and

![Figure 3. V\(H\) and V\(\kappa\) gene positions relative to CTCF occupancy.](image-url)

(A) Overview of V\(H\) gene positions relative to the nearest CTCF binding site for the mouse Igh locus. The germline Igh locus is shown with the locations of V\(H\), D\(H\), J\(H\), C\(H\), and enhancer elements (top). V\(H\) genes were grouped according to the location of the nearest CTCF site and whether this site was located upstream (−) or downstream (+) relative to V\(H\) transcripitinal orientation. Color-coded groups of V\(H\) genes were plotted at their respective locations below the germline V\(H\) genes. On the right, a pie-chart represents the distribution of the different V\(H\) gene groups expressed in percentages of the total number of V\(H\) genes. (B) Identical to panel A, but for the V\(\kappa\) genes within the mouse Igk locus.

From www.bloodjournal.org by guest on April 1, 2019. For personal use only.
Figure 4. CTCF distribution and chromosomal breakpoint regions (BPR) around Ig enhancers and oncogenes involved in translocation with antigen receptor loci. (A) Genomic regions surrounding the human IgH enhancers. CTCF ChIP-Seq data (from the human GM12878 B lymphoblastoid cell line) is shown above the localization of D, H, and C fragments. Green represents enhancers; and cyan, most common chromosomal BPRs. For the IgH locus, 2 common BPRs are annotated. eBL indicates endemic Burkitt lymphoma; sBL, sporadic Burkitt lymphoma; and HIV BL, HIV-induced Burkitt lymphoma. For scaling purposes, the position of E (IGH), Ei (IGK), or HuE (IGL) enhancers was set to 0. (B) CTCF ChIP-Seq data (from the human GM12878 B lymphoblastoid cell line) are shown for 3 genomic regions around oncogenes involved in translocations with antigen receptor loci. Locations of genes are shown on top, with the relevant oncogenes in blue and other genes in gray. The oncogene transcriptional start site was set to 0 for scaling purposes. Cyan indicates the most common chromosomal BPRs. BPRs linked to distinct B-cell malignancies are indicated as in panel A. For MYC, breakpoints found specifically in translocations with Ig light chain loci are annotated (IGK/GL). Red bars represent the BCL6 CpG islands located in intron 1.

are present in the same transcription factories. Recently developed new genome-wide methods that measure primary chromosomal rearrangements in the absence of growth selection have shown that translocations between IGH and MYC and all other genes are indeed directly related to their contact frequencies. However, 3C-based measurements showed that approximately 30% of all genes interact with IGH at equal or higher frequency than MYC. Therefore, growth selection and AID-mediated DNA damage rather than a high contact frequency between IGH and the translocations partners account for their high rate of translocation.

An oncogenic role of the IGH locus 3′RR has been defined based on long-range activation of translocated MYC genes. In this context, an important function of CTCF is to restrict enhancer interaction to Ig loci (Figure 4). Obviously, CTCF-mediated restriction might be lost when strong Ig enhancers appear in an aberrant genomic context in chromosomal translocations, which may then lead to activation of oncogenes. In addition, CTCF sites clustered around Ig enhancers may affect local chromatin conformation to promote novel promoter-enhancer interactions. The genes encoding c-Myb and c-Myc, 2 essential hematopoietic regulators and oncogenic transcription factors, harbor multiple CTCF-binding sites in their vicinity (Figure 4B). Indeed, CTCF was originally discovered as a transcriptional repressor of c-Myc. Both genes are regulated at the level of transcriptional elongation by attenuator elements located in their first introns, close to CTCF-binding sites (Figure 4B). The exact role played by CTCF at these sites is still unclear, but the recent demonstration that CTCF regulates RNA polymerase II processing suggests that CTCF could directly affect elongation. Loss of intronic CTCF binding resulting from translocations could disrupt normal locus structure and transcriptional control, allowing the juxtaposed IGH enhancers to drive MYC expression. Because MYB seems to be regulated by a similar mechanism, this may also occur within the MYB locus, when translocated to the TCR-β locus in acute T-cell leukemia.

Interestingly, based on published CTCF-binding sites, translocation breakpoints at the CRLF2 and BCL3 loci found in several types of B-cell malignancies are both confined to an approximately 25-kb region between the transcriptional start site and an upstream CTCF element. Translocations therefore result in loss of the upstream CTCF-binding site, which may allow strong Ig enhancers to interact and activate the CRLF2 and BCL3 promoters. In contrast, many other loci frequently involved in leukemic translocations (e.g., BCL2 and MYC), show more complex breakpoint patterns and harbor numerous CTCF-binding sites (Figure 4B).

Expression of BCL6 in diffuse large B cell lymphomas is maintained through hypermethylation of intragenic CpG islands in
its first intron.\(^{66}\) Hypermethylated, which has been observed in a wide range of cancer cells, prevents CTCF-mediated silencing of BCL6 because CTCF binds these CpG islands in a methylation-sensitive fashion. Furthermore, long-range chromosome regions approximately 150 to 260 kb upstream have been implicated in BCL6 regulation,\(^{57}\) possibly representing another layer of control in which CTCF might play an important role (Figure 4B).

In conclusion, based on studies in the Igk and Igx, it has become clear that CTCF-binding provides enhancer-blocking activity that restricts Ig enhancer activity to the Ig loci. Nevertheless, CTCF also serves to bring remote V gene promoters into close proximity of Ig enhancers. It will be challenging to identify the exact mechanism by which CTCF regulates V gene choice because individual developing B cells are expected to manifest distinct and highly dynamic interactions between V regions and Ig enhancers. Moreover, the high density of CTCF-binding sites and their anticipated partial redundancy may complicate the interpretation of in vivo targeted deletion experiments.

Chromosomal translocations present in lymphoid malignancies will (1) remove CTCF sites close to the oncogenes, (2) bring in enhancer-associated CTCF sites, or (3) dramatically alter the spacing between CTCF sites and their cognate genes. These events are likely to influence local communication between cis-regulatory elements by imposing new structural constraints and thereby increasing oncogene expression. Additional epigenetic changes, including DNA methylation and histone modifications, might also interfere with CTCF function. For example, the BCL6 oncogene is involved in chromosomal translocations through breakpoints frequently found near the intronic CpG islands where CTCF binds in a methylation-sensitive manner.\(^{65,68}\)

In a normal context, oncogene expression is tightly regulated in response to extracellular signaling events. It is an intriguing question why these signals are no longer able to efficiently regulate expression of the translocation products. Trying to address this issue might reveal additional roles for CTCF in the integration of signaling to chromatin. Future research should clarify the complex roles played by CTCF in the wiring of cis-DNA regulatory elements and in organizing networks of genomic interactions. The field of spatial chromatin organization research is currently exploding with the development of new exciting technologies, allowing high-throughput analysis of chromatin organization of the whole genome.\(^{27,89,90}\) As the resolution of these technologies increases, they might rapidly become valuable tools to reveal the complex 3-dimensional changes in cancer genomes and unravel their relationships with critical regulatory factors, including CTCF.

Acknowledgments

This work was supported in part by Fundação para a Ciência e a Tecnologia (C.R.d.A.), Royal Netherlands Academy of Arts and Sciences (R.S.), the Center of Biomedical Genetics, the Cancer Genomics Center, and the EuTRACC Consortium (E.S. and R.S.).

Authorship

Contribution: C.R.d.A. wrote the manuscript and designed figures; R.S. analyzed and interpreted data, wrote the manuscript, and designed figures; S.T. analyzed data; E.S. wrote the manuscript; and R.W.H. designed figures and wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

The current affiliation for C.R.d.A. is Genome Function Group, Medical Research Council Clinical Sciences Centre, Imperial College School of Medicine, London, United Kingdom.

Correspondence: Rudi W. Hendriks, Department of Pulmonary Medicine, Room Ez2251a, Erasmus MC Rotterdam, PO Box 2040, NL 3000 CA Rotterdam, The Netherlands; e-mail: r.hendriks@erasmusmc.nl.

References


51. Degner SC, Wong TP, Jankevicius G, Feeney AJ. Cutting edge: developmental stage-specific re-


DNA-binding factor CTCF and long-range gene interactions in V(D)J recombination and oncogene activation

Claudia Ribeiro de Almeida, Ralph Stadhouders, Supat Thongjuea, Eric Soler and Rudi W. Hendriks