

Published in final edited form as:

Adv Cancer Res. 2011 ; 110: 43–76. doi:10.1016/B978-0-12-386469-7.00003-7.

Chromatin Insulators: A Role in Nuclear Organization and Gene Expression

Jingping Yang and Victor G. Corces

Department of Biology, Emory University, Atlanta, GA 30322, USA

Jingping Yang: jyang29@emory.edu

Abstract

Chromatin insulators are DNA-protein complexes with broad functions in nuclear biology. Based on the ability of insulator proteins to interact with each other, it was originally thought that insulators form loops that could constitute functional domains of co-regulated gene expression. Nevertheless, data from genome-wide localization studies indicate that insulator proteins can be present in intergenic regions as well as at the 5', introns or 3' of genes, suggesting a broader role in chromosome biology. Cells have developed mechanisms to control insulator activity by recruiting specialized proteins or by covalent modification of core components. Recent results suggest that insulators mediate intra- and inter-chromosomal interactions to affect transcription, imprinting and recombination. It is possible that these interactions set up cell-specific blueprints of nuclear organization that may contribute to the establishment of different patterns of gene expression during cell differentiation. As a consequence, disruption of insulator activity could result in the development of cancer or other disease states.

INTRODUCTION

Patterns of transcription required for cell differentiation are initially established by specific transcription factors. The maintenance of these patterns of gene expression is then carried out by alterations in chromatin structure that are epigenetically inherited between cell generations. These changes in chromatin organization take place at the level of the 10 nm fibre and include covalent histone modifications, DNA methylation and alterations induced by ATP-dependent remodelling complexes. In addition, recent evidence suggests that the three-dimensional organization of the genome within the nucleus of eukaryotic cells may also be critical for achieving proper spatial and temporal patterns of gene expression during development. The factors and processes involved in the establishment, maintenance and regulation of specific states of nuclear organization are largely unknown but insulators are emerging as likely candidates to play this crucial role (Bushey, Dorman et al. 2008; Phillips and Corces 2009).

Insulators are DNA-protein complexes experimentally defined by their ability to block enhancer-promoter interactions and/or serve as barriers against the spreading of the silencing effects of heterochromatin. Not all sequences described as insulators display these two properties; Felsenfeld and colleagues have proposed to classify these sequences as “enhancer blocking” or “barrier” insulators depending on whether they interfere with enhancer promoter interactions or antagonize spreading of heterochromatin under experimental conditions (Gaszner and Felsenfeld 2006). Barrier insulators appear to function by recruiting enzymes involved in histone modifications that stop the spreading of silencing

marks (Huang, Li et al. 2007; Dhillon, Raab et al. 2009) and will not be considered in this review. Instead, we concentrate our discussion on the function of enhancer blocking insulators with special emphasis on recent data supporting a role in nuclear organization. These insulators were originally discovered in *Drosophila*, which has been a particularly good model system in which to analyze insulator function. Several different insulators have been identified in this organism whereas vertebrates appear to mostly rely on the CTCF insulator (Phillips and Corces 2009). We conclude by proposing that the primary role of insulators may not be to regulate enhancer-promoter interactions or heterochromatin spreading. Rather, insulators may mediate intra- and inter-chromosomal interactions with the primary goal of organizing the eukaryotic genome into epigenetically heritable states. This insulator-mediated organization may be important to regulate DNA function at multiple levels, including transcription initiation, elongation and DNA recombination. Miss-regulation of insulator function may lead to alterations in gene expression and the development of disease states and cancer.

INSULATOR STRUCTURE: CORE COMPONENTS

There are several types of insulators in *Drosophila* that have been studied in detail. They include the *scs* and *scs'* sequences originally discovered flanking the heat shock *hsp70* locus (Kellum and Schedl 1992; Zhao, Hart et al. 1995), the *gypsy* insulator first found in the *gypsy* retrotransposon (Geyer and Corces 1992; Hoover, Chien et al. 1993), the Fab-7/Fab-8/Mcp insulators located in the bithorax complex (Gyurkovics, Gausz et al. 1990; Karch, Galloni et al. 1994) and the SF1 insulator described in the Antennapedia complex (Belozero, Majumder et al. 2003). Each of these insulators consists of a DNA sequence and a specific DNA binding protein that interacts with this sequence; additional factors that interact with the DNA binding protein have been characterized for some of these insulators (Figure 1).

In the case of the *scs* insulator, the DNA binding protein component is Zeste-White 5 (ZW5), which is a Zn finger protein required for cell viability. Null mutations in the *zw5* gene are recessive lethal, but hypomorphic alleles display a variety of pleiotropic effects on wing, bristle and eye development (Gaszner, Vazquez et al. 1999). The *scs'* sequences interact with a protein called Boundary Element Associated Factor 32 (BEAF 32). The *BEAF 32* gene encodes two different proteins named BEAF 32A and BEAF 32B, which are present at hundreds of sites on *Drosophila* polytene chromosomes (Zhao, Hart et al. 1995). The two isoforms differ at the N terminal DNA binding domain (BED Finger domain). The common C terminal region is involved in protein-protein interactions between the two isoforms. Analysis of mutations in the *BEAF 32* gene shows that BEAF 32B is required for viability whereas BEAF 32A mutations do not show significant phenotypic defects. Expression of a dominant negative form of BEAF 32 results in changes in chromosome structure and cell viability (Gilbert, Tan et al. 2006; Roy, Gilbert et al. 2007; Roy, Tan et al. 2007).

The *gypsy* insulator contains binding sites for the Suppressor of Hairless-wing [Su(Hw)] protein, which is an 12 zinc finger DNA binding protein. Mutations in the *su(Hw)* gene cause female sterility but do not result in lethality (Harrison, Gdula et al. 1993). Su(Hw) interacts with two other components of the *gypsy* insulator, Mod(mdg4)2.2 and CP190 (Gause, Morcillo et al. 2001; Ghosh, Gerasimova et al. 2001). Mod(mdg4)2.2 does not bind to DNA directly but interacts with Su (Hw) through its carboxy-terminal domain. In addition, Mod(mdg4)2.2 also contains a BTB domain in the N-terminal domain that mediates homo and hetero-multimerization with other insulator components. The *mod(mdg4)* gene encodes approximately 29 different isoforms that arise by alternative *cis*- and *trans*-splicing (Buchner, Roth et al. 2000; Labrador, Mongelard et al. 2001; Labrador

and Corces 2003); null mutations in the gene result in lethality, but mutations affecting the Mod(mdg4)2.2 isoform are viable and show defects in *gypsy* insulator function (Gerasimova, Gdula et al. 1995). CP190 also contains a BTB domain as well as 3 zinc fingers and it interacts with both Su(Hw) and Mod(mdg4)2.2. The BTB domains of Mod(mdg4)2.2 and CP190 also mediate homo di- and multi-merization of these two proteins. CP190 binds DNA with low affinity and specificity but it does not interact directly with insulator sequences present in the *gypsy* retrotransposon, where it is recruited through interactions with Su(Hw) and Mod(mdg4)2.2 instead. Mutations in the CP190 gene are lethal (Pai, Lei et al. 2004).

The bithorax complex of *Drosophila* contains an intricate collection of transcriptional regulatory sequences that orchestrate the complex spatio-temporal expression of the three genes present in the complex. The proper interplay between these regulatory sequences requires the function of several insulators of which Fab-8 has been studied in most detail. Fab-8 sequences interact with the *Drosophila* homolog of the vertebrate CTCF insulator protein. *Drosophila* CTCF (dCTCF) has 12 zinc fingers. Mutations in *dCTCF* are lethal and show abdominal homeotic phenotypes (Gerasimova, Lei et al. 2007; Mohan, Bartkuhn et al. 2007). dCTCF is also found in the Mcp and Fab-6 insulators present in the bithorax complex but not Fab-7 (Holohan, Kwong et al. 2007). Fab-7 may represent a fourth class of insulators that use the GAGA factor (GAF) as a DNA binding protein that also contains a BTB domain. Mutations in the *trl* gene, which encodes GAF, affect Fab-7 insulator activity (Schweinsberg, Hagstrom et al. 2004). In addition, GAF is present and required for the function of the SF1 insulator found in the Antennapedia complex (Belozarov, Majumder et al. 2003).

If the functions of all *Drosophila* insulators have converged into that of CTCF in vertebrates, one may expect to find some shared protein components among *Drosophila* insulators. This is indeed the case. The CP190 protein, first found in the *gypsy*/Su(Hw) insulator, also interacts with dCTCF (Gerasimova, Lei et al. 2007; Mohan, Bartkuhn et al. 2007). Genome-wide mapping of dCTCF and CP190 sites also support this conclusion (Bartkuhn, Straub et al. 2009; Bushey, Ramos et al. 2009). These studies have also shown that BEAF and CP190 co-localize at hundreds of sites throughout the genome (Bushey, Ramos et al. 2009). These results suggest that the insulators defined by these three different DNA binding proteins, Su(Hw), dCTCF and BEAF share the BTB domain-containing protein CP190 and may therefore use similar mechanisms to effect their insulator function. On the other hand, GAF does not appear to interact directly with CP190 but has been shown to interact with Su(Hw) and Mod(mdg4)2.2 (Melnikova, Juge et al. 2004); since these two proteins can in turn interact with CP190, GAF insulators may act mechanistically like the other three types (Figure 1).

Vertebrates appear to rely mostly on the widespread CTCF insulator (Wallace and Felsenfeld 2007) and we will concentrate our discussion on this sequence. CTCF is a highly conserved protein containing an eleven zinc finger central DNA binding domain, displaying close to 100% homology between mouse, chicken, and human, embedded within slightly more divergent N- and C-termini (Ohlsson, Renkawitz et al. 2001). On the basis of its ability to bind to a wide range of variant sequences as well as specific co-regulatory proteins through combinatorial use of different zinc fingers, CTCF was originally described as a multivalent factor (Filippova, Fagerlie et al. 1996). This structural feature may explain the ability of vertebrate cells to fulfil all insulator functions with just one insulator whereas *Drosophila* cells require multiple insulators. CTCF homozygous knockout mice exhibit early embryonic lethality prior to implantation, highlighting the critical importance of CTCF in diverse cellular processes (Splinter, Heath et al. 2006; Heath, Ribeiro de Almeida et al. 2008).

Interestingly, the Mod(mdg4) and CP190 proteins have not been conserved in vertebrates. Instead, the vertebrate CTCF protein interacts with cohesin, which has recently emerged as a critical partner of CTCF in mediating inter- and intra-chromosomal interactions necessary for transcription and insulator function. Cohesin is a ring-shaped complex that holds chromatids together between the S and M phases of the cell cycle. In vertebrates, cohesins are present at CTCF target sequences in the genome and this localization depends on CTCF (Parelho, Hadjur et al. 2008; Rubio, Reiss et al. 2008; Stedman, Kang et al. 2008; Wendt, Yoshida et al. 2008). Cohesin is necessary for nearly all the intra-chromosomal interactions attributed to CTCF, such as those observed at the *H19/IGF2*, *β -globin*, *IFNG* and *APO* loci; knockdown of cohesin impairs chromatin interactions at these loci (Hadjur, Williams et al. 2009; Mishiro, Ishihara et al. 2009; Nativio, Wendt et al. 2009; Hou, Dale et al. 2010). These observations have led to the proposal that the ring-like structure of the cohesin complex can mediate interactions between distant genomic sites by a mechanism similar to that employed to maintain sister chromatids together.

INSULATOR STRUCTURE: ACCESSORY FACTORS

CTCF has been found to interact with other nuclear factors in addition to cohesins, although the broad relevance of some of these interactions to general aspects of CTCF function is unclear. For example, CTCF interacts *in vivo* with CHD8, which is a member of the chromodomain helicase (CHD) family and has an SNF2-like helicase/ATPase domain. CHD8 has been shown to be present at some CTCF target sites, including the *H19/IGF2* Imprinting Control Region (ICR), the promoter regions of the *BRCA1* and *C-MYC* genes, and the 5'HS5 insulator of the *β -globin* locus in human cells (Ishihara, Oshimura et al. 2006). Like CTCF, CHD8 is also required for the insulator activity of the *H19/IGF2* ICR. Loss of CHD8 leads to expression of *IGF2* from the maternal and paternal alleles (Ishihara, Oshimura et al. 2006). CHD8 together with CTCF can affect aspects of chromatin structure such as CpG DNA methylation and histone acetylation. Knockdown of CHD8 leads to hypermethylation of CpGs at CTCF binding sites in the promoter regions of the *BRCA1* and *C-MYC* genes (Ishihara, Oshimura et al. 2006). In addition to CHD8, CTCF can also interact *in vitro* with SIN3A which recruits a histone deacetylase activity necessary for the silencing function of CTCF (Lutz, Burke et al. 2000); this result has not been confirmed by *in vivo* experiments. These observations suggest that CTCF may have context-dependent functions that are mediated by different protein partners.

CTCF has been also found to interact with transcription factors such as YY1, which is a zinc finger protein capable of activating or repressing transcription depending on the promoter context. Like CTCF, YY1 also binds to DNA in a methylation-sensitive manner. YY1 can bind the *PEG3* ICR and regulate its parent of origin-dependent expression. Insulator assays with deletion constructs of YY1-binding sites indicate that the region functions as a methylation-sensitive insulator like the *H19/IGF2* ICR (Kim, Kollhoff et al. 2003). YY1 is a required cofactor for CTCF in processes such as X chromosome inactivation. Both YY1 and CTCF bind at *Tsix* DNA, and a deficiency of YY1 leads to aberrant *Tsix* and *Xist* expression, aberrant XCI patterns, and results in a deficit of male and female embryos. YY1 and CTCF together function as *Tsix* transcriptional activators in ES cells (Donohoe, Zhang et al. 2007). It is not yet clear whether YY1 alone can act as an insulator or only contributes to the insulator activity of CTCF as a cofactor. In addition to YY1, CTCF can interact with other transcription factors such as YB1, Oct4, Kaiso and thyroid hormone receptor (TR). CTCF interacts with YB1 *in vitro* and *in vivo* and functions as a co-repressor at the *C-MYC* promoter (Chernukhin, Shamsuddin et al. 2000). CTCF interacts with Kaiso *in vivo* and both bind to the chicken HS4 insulator. Kaiso inhibits the blocking activity of CTCF in transgenic assays (Defossez, Kelly et al. 2005). CTCF interacts with Oct4 and control X-chromosome pairing during X-chromosome inactivation (Donohoe, Silva et al. 2009). A

subset of CTCF binding sites is found next to TR response elements in the mouse *C-MYC* and the human *APP* genes. Functional TR binding is required for enhancer blocking activity (Weth, Weth et al. 2010). Intriguingly, CTCF has been also found to interact with RNA polymerase II. The largest subunit of Pol II associates with CTCF target sites *in vivo* and this interaction is dependent on the presence of intact CTCF binding sequences. A single CTCF binding site in the *C-MYC* gene (N-MYC-CTCF), but not a mutant site deficient for CTCF binding, is sufficient to activate transcription from a promoterless reporter gene in stably transfected cells (Chernukhin, Shamsuddin et al. 2007). This suggests that CTCF may recruit Pol II and activate transcription. The extent of such a role for CTCF is unclear, since genome-wide studies of CTCF sites suggest that only a subset of CTCF sites are located at promoters in the human genome (Kim, Abdullaev et al. 2007).

The function of *Drosophila* and vertebrate insulators appears to rely in part on their ability to interact with the nuclear lamina. *Drosophila* insulator proteins fractionate with components of the nuclear lamina, suggesting a direct or indirect interaction between the two (Byrd and Corces 2003). Furthermore, the dTopors protein, which is located in the nuclear periphery, interacts with both Mod(mdg4)2.2 and Lamin, and mutations in the *Lamin Dm0* gene results in disruption of insulator activity (Capelson and Corces 2005). It appears that interaction with a nuclear substrate is also required for the proper function of the CTCF insulator. For example, it has been shown that CTCF interacts with nucleophosmin, a nuclear matrix protein that is concentrated at the surface of the nucleolus and associates with CTCF at the chicken HS4 insulator in the β -globin locus. Consistent with this distribution, transgenic copies of reporter genes containing the chicken HS4 insulator show preferential localization to the outer part of the nucleolus, and this localization depends on the presence of an intact CTCF binding site (Yusufzai, Tagami et al. 2004). In addition to nucleophosmin, other nuclear matrix proteins such as Lamin may also be important for CTCF function. In human cells, many of the Lamin B1-associated domains (LADs) in the genome are flanked by CTCF (Guelen, Pagie et al. 2008). Lamin A also works together with CTCF to establish nucleolar localization. The D4Z4 human subtelomeric repeats localize telomeres to the nuclear periphery and display insulator activity in transgenic cell lines. CTCF and Lamin A both bind to D4Z4 and are required for proper perinuclear localization. Reduction of either protein results in a decrease in the localization of telomeres at the nuclear periphery and impairs insulator function (Ottaviani, Rival-Gervier et al. 2009; Ottaviani, Schluth-Bolard et al. 2009). It is thus possible that the interaction between CTCF and other nuclear matrix proteins contributes to chromatin localization and organization in the nucleus in a variety of cell processes.

These results suggest that *Drosophila* and vertebrate insulators may play a variety of functions in nuclear biology. Many of these functions can be explained by the ability of CTCF and *Drosophila* insulator proteins to mediate interactions between two or more chromosomal locations. Different *Drosophila* insulators may have specialized roles in effecting distinct nuclear functions based on the presence of particular protein components. In vertebrates, it is not clear whether CTCF acts by different mechanisms that are context dependent and vary based on its association with the different partners described above. It is nevertheless possible that these different functions have a common requirement for the ability of CTCF to mediate interactions between distant sites in the genome; this may be also the case in some of the more classical roles characteristic of standard transcription activators and repressors. Although the DNA recognition sequence of CTCF appears to be conserved at different genomic locations (Kim, Abdullaev et al. 2007), the distinct context-dependent roles of this protein must be mediated by the different protein partners with which it associates. The nature of the determinants responsible for these context-dependent interactions is unknown and should be an important issue for future investigation.

GENOMIC DISTRIBUTION OF INSULATORS

The existence of several *Drosophila* insulator subclasses with different DNA binding proteins but sharing some functional components raises the question of whether they all have the same role in the regulation of gene expression or whether there is a functional specialization in their tasks. The possibility of such specialization is highlighted by results showing differential localization of insulator subclasses with respect to genomic landmarks. For example, Su(Hw) and dCTCF are preferentially excluded from exonic regions (mostly 5' and 3' UTRs), with only 8%, 16%, and 17% of sites found within exons, respectively, whereas BEAF sites are enriched in UTRs (Bushey, Ramos et al. 2009). When the location of these proteins is compared with the location of genes, few Su(Hw)-binding sites are found in the 1 kb regions flanking genes. However, dCTCF and BEAF show a distribution that is highly skewed toward the 5' end of genes and is enriched in the first 200 bp just upstream of the transcription start site. Insulator proteins also show a compartmentalized distribution in relation to the level of gene expression. For example, 83% of dCTCF sites and 89% of BEAF sites at the 5' end of genes localize to genes that are highly expressed. However, Su(Hw)-binding sites are most often found near genes with low expression levels. Finally, different insulator proteins appear to associate with genes involved in different cellular processes. Genes containing dCTCF in the 200 bp region upstream of their transcription start site are mostly involved in developmental processes, whereas genes containing BEAF in this region are mostly involved in metabolic processes. Both dCTCF and BEAF are enriched near or at genes involved in cell cycle processes, whereas Su(Hw)-containing genes show little significant clustering based on biological process (Emberly, Blattes et al. 2008; Bushey, Ramos et al. 2009; Jiang, Emberly et al. 2009). These observations suggest a division of labour among *Drosophila* insulators, both with respect to gene function as well as specific aspects of cell function. Given the fact that all insulators share CP190 and perhaps Mod(mdg4), it is likely that all of them use the same mechanism to perform their function, namely bring together different regions of the genome. Nevertheless, *Drosophila* cells appear to use a variety of DNA binding insulator proteins to recruit other insulator components to mediate these interactions. Therefore, the specific outcome of these interactions may be determined by where in the genome the binding sites for each of these proteins are localized. In the case of Su(Hw) and a subset of dCTCF sites, their localization in intergenic regions suggests that their role may be to form loops that may represent independent functional domains. The rest of dCTCF sites and BEAF sites are located around promoter regions and their function may rely on the same type of interactions to bring these regions of genes to specific nuclear compartments such as transcription factories.

Since *Drosophila* has several insulators with distinct localization patterns with respect to genomic features, it is interesting to contrast this information with the genome wide localization of CTCF in vertebrates. The distribution of CTCF in mouse embryonic stem cells and in various human cell lines has been studied using ChIP-chip and ChIP-seq approaches (Barski, Cuddapah et al. 2007; Kim, Abdullaev et al. 2007; Chen, Xu et al. 2008; Jothi, Cuddapah et al. 2008; Cuddapah, Jothi et al. 2009). The number of sites uncovered varies depending on the approach, from ca. 13,800 in IMR90 human fibroblasts using ChIP-chip (Kim, Abdullaev et al. 2007) to ca. 39,600 in mouse ES cells using ChIP-seq (Chen, Xu et al. 2008). In general, the distribution of CTCF sites on each chromosome correlates with gene density. However, different from transcription factors, CTCF binding sites are generally located an average of 48 kb away from promoters. Approximately 46% of the sites are located in intergenic regions, consistent with the classical role of insulators in establishing chromatin domains, whereas ca. 20% of the sites display promoter proximal localization. As in *Drosophila*, a significant number of the sites fall within genes, with 22% in introns and 12% in exons in human fibroblast cells (Kim, Abdullaev et al. 2007). Analyses in resting human CD4⁺ T cells have found a similar genomic distribution of CTCF

sites, with 45% intergenic, 7% 5'UTR, 3% exonic, 29% intronic, 2% 3'UTR, and 13% within 5 kb of the transcription start site. Although CTCF sites generally correlate with gene density, it is possible to find patterns in their distribution, other than location with respect to genes, suggestive of various roles in gene expression. For example, CTCF depleted domains, which exhibit lower-than-average CTCF density, tend to include clusters of related gene families and genes that are transcriptionally coregulated. In addition, some CTCF sites are located in intergenic regions at the transition of chromatin domains with different epigenetic status. For example, there are 793 CTCF sites flanking H3K27me3 domains in CD4⁺ T cells and this occupancy may be cell type specific (Cuddapah, Jothi et al. 2009). CTCF can also bind 5–10 kb outside Lamina-associated domains (LADs), which contain low gene density and low expression genes, representing a repressive chromatin environment; out of 1344 LADs found, 333 have CTCF binding at least at one side of the border (Guelen, Pagie et al. 2008). In all three cases, the distribution of CTCF agrees with a classical insulator role for this protein in the establishment or maintenance of functional transcriptional domains. On the other hand, CTCF enriched domains, which have higher than average CTCF binding, often have multiple alternative promoters (Kim, Abdullaev et al. 2007). These and other sites located close or within genes may play roles in chromosome biology different from those classically assigned to insulators.

The genome occupancy of CTCF sites is similar in different cell types but a significant fraction appears to be cell-type specific. For example, 40–60% of CTCF sites are common among CD4⁺ T, HeLa and Jurkat cells (Cuddapah, Jothi et al. 2009) and the rest of the sites occupied by CTCF vary between the different cell types. Similarly, around 30% of CTCF sites are different in human IMR90 fibroblasts and U937 erythroid progenitor cells (Kim, Abdullaev et al. 2007). Nevertheless, as is the case in *Drosophila*, a number of CTCF sites in mammals are cell type-specific and may play a role in establishing patterns of gene expression required for cell differentiation.

INSULATORS MEDIATE INTRA- AND INTER-CHROMOSOMAL INTERACTIONS

All *Drosophila* insulators, with perhaps the exception of *scs*/ZW5 for which data is unavailable, share the BTB domain proteins CP190 and Mod(mdg4), although the specific isoform of Mod(mdg4) appears to be different in different insulators. The BTB domains of these two proteins as well as GAF can interact with each other in various *in vitro* or *in vivo* assays, suggesting that insulator proteins may mediate intra- and inter-chromosomal interactions among insulator sites throughout the genome. Various types of observations support this conclusion. For example, Su(Hw), Mod(mdg4)2.2, dCTCF and CP190 show a punctuate distribution in the nuclei of diploid cells. These sites, called insulator bodies, appear to contain multiple individual insulator sequences and their morphology is disrupted by mutations in insulator components (Gerasimova, Byrd et al. 2000; Pai, Lei et al. 2004; Gerasimova, Lei et al. 2007). Furthermore, FISH experiments have shown that DNA sequences contained between two insulators form a loop, which becomes two smaller loops when a new insulator is inserted in the middle of the DNA (Byrd and Corces 2003). Although *Drosophila* insulator bodies are present throughout the nucleus, they seem to localize preferentially in the nuclear periphery. This localization may be mediated by the protein dTopors, which has been shown to interact with Su(Hw) and Mod(mdg4)2.2 as well as with Lamin Dm0 (Capelson and Corces 2005; Capelson and Corces 2006). Therefore, dTopors may serve as an anchor to attach insulator sites to the nuclear lamina or nuclear matrix (Capelson and Corces 2005). In addition to this type of information based on FISH and immunolocalization experiments, results from 3C experiments have shown that *scs* and *scs'* sequences are present in close proximity in the nucleus forming a loop of the intervening sequence. These contacts may be mediated by CP190 or may be direct, since

ZW5 and BEAF 32 have been shown to interact both *in vitro* and *in vivo* in *Drosophila* embryos (Blanton, Gaszner et al. 2003).

The role of insulators in mediating intra- and inter-chromosomal interactions has been better established in vertebrates using 3C, 4C and FISH approaches in a number of different systems. The β -globin locus contains several CTCF sites that are conserved in mice and humans. These sites are located within two DNase I hypersensitive regions that flank the β -globin locus: the 5'HS5 located in the locus control region (LCR) and the downstream 3'HS1; additional CTCF sites are located further away at either end of the locus. The 5'HS5 and 3'HS1 sites interact with each other as well as with the more distal CTCF sites flanking the locus and these interactions are severely weakened in the absence of the CTCF protein. The interactions with 5'HS5 or 3'HS1 are cell type specific but they are not required for proper transcription of the β -globin gene. These interactions exist in both erythroid cells, in which the β -globin gene is transcribed and in erythroid progenitor and fibroblast cells in which the gene is not expressed, but not in non-expressing brain cells (Tolhuis, Palstra et al. 2002; Splinter, Heath et al. 2006; Hou, Dale et al. 2010). Mutation of 3'HS1 destabilizes the long-range interactions among these CTCF sites but mutation or deletion of 3'HS1 or 5'HS5 can neither affect the expression kinetics nor levels of the β -globin genes in erythroid cells (Farrell, Grinberg et al. 2000; Splinter, Heath et al. 2006). Nevertheless, a number of CTCF sites present outside of the β -globin locus interact with each other but not with 5'HS5 or 3'HS1, and some of these interactions are both cell type specific and transcription relevant.

These interactions can only be detected in either erythroid cells or non-erythroid cells. Knockdown of CTCF, which leads to a global reduction of both groups of interactions, can negatively affect β -globin gene transcription (Hou, Dale et al. 2010). It is possible that these interactions are required for β -globin expression, although it is also possible that the observed changes in transcription are due to indirect effects of CTCF knockdown. Therefore, it appears that CTCF-mediated interactions involving 5'HS5, 3'HS1 and other CTCF sites create a three-dimensional organization of the β -globin locus before the commitment to the erythroid lineage that is not necessary or sufficient for the activation of the globin genes. The fact that CTCF-mediated interactions among these sites are not observed in brain cells suggests that this organization may play an earlier role during differentiation after the commitment to ectodermal or mesodermal fates. New interactions established later during erythroid differentiation and involving a different set or more distally located CTCF sites may be involved in the establishment of a more global architecture that is actually more directly responsible of the expression of the β -globin locus.

The requirement of CTCF-mediated interactions for gene expression has been studied in several other loci. At the cytokine interferon- γ (*IFNG*) locus there are three conserved CTCF binding sites located upstream, downstream and within the gene 1.5 kb from the transcription start site (TSS) in human and mouse cells. The CTCF sites display cell type-specific interactions in specialized T helper 1 (Th1) cells and these interactions parallel differences in CTCF occupancy (Hadjur, Williams et al. 2009; Sekimata, Pérez-Melgosa et al. 2009). These CTCF sites also interact with the enhancers present in the locus, although the enhancers do not have CTCF binding sites. The three-dimensional conformation of the locus and robust *Ifng* expression in Th1 cells are dependent on the presence of CTCF (Sekimata, Pérez-Melgosa et al. 2009). It is possible that interactions among the CTCF sites help recruit the enhancers and activate gene expression at the locus. The apolipoprotein (APO) gene (APOA1/C3/A4/A5) cluster in humans is another interesting example of CTCF-mediated loops that are required for proper gene expression. Results from 3C experiments indicate that CTCF sites in this region interact with each other in Hep3B cells. These interactions lead to the formation of two transcribed loops with APOC3/A4/A5 in one loop and APOA1 in the other. Consistent with the model based on 3C data, CTCF knockdown

leads to a decrease of APOC3/A4/A5 but an increase of APOA1 expression (Mishiro, Ishihara et al. 2009). Major histocompatibility class II (MHC-II) genes are also regulated by intra-chromosome interactions mediated by CTCF. The XL9 element contains CTCF sites and is located in the intergenic region between the MHC-II genes HLA-DRB1 and HLA-DQA1. Knockdown of CTCF diminishes expression of these genes. 3C experiments detect an interaction between XL9 and proximal promoter elements of these two MHC-II genes. These interactions also depend on the class II transactivator (CIITA), which can associate with CTCF (Majumder, Gomez et al. 2008). The chromatin conformation mediated by the associated between CIITA and CTCF is required for transcription of the MHC-II genes. In addition to its role in processes that result in increased transcription of genes, CTCF may also mediate chromatin high order structure that results in gene silencing. For example, 3C experiment suggest that silenced Hox genes in human cells are spatially clustered via a specific three-dimensional architecture of the locus; CTCF is present at the sites of contact and appears to be responsible for the establishment and/or maintenance of this architecture (Ferrauiolo, Rousseau et al. 2010).

The three-dimensional arrangement of the chromatin fibre created by CTCF-mediated interactions also plays an important role in imprinted gene expression at the *H19/IGF2* locus. The imprinting control region (ICR) immediately upstream of *H19* contains CTCF sites that are crucial for the imprinting of these two genes. Results from 3C experiments demonstrate that these CTCF sites can mediate allele-specific chromosome interactions that seem to control the accessibility of the *IGF2* promoter to the shared enhancer. On the maternal allele of both mouse and human cells the ICR is not methylated and CTCF binds to DNA. In the mouse, the ICR interacts with CTCF sites present in the upstream DMR1 and downstream MAR3 sites that flank the *IGF2* gene. The three-dimensional structure formed as a consequence of these interactions keeps *IGF2* in an enclosed domain (Murrell, Heeson et al. 2004; Kurukuti, Tiwari et al. 2006). In human cells, the ICR interacts with a CTCF Downstream Site (CTCF DS) located downstream of the shared enhancer; this interaction creates a loop that encloses the enhancer (Nativio, Wendt et al. 2009). In both mouse and human maternal chromosomes the enhancer is unable to interact with the *IGF2* promoter. In mouse, the 142* allele has a mutated CTCF binding site in the ICR, and in mice that inherit this allele from their mother, the ICR interacts with DMR2 instead of DMR1 and MAR3; this is accompanied by biallelic expression of *IGF2*. On the other hand, there is no effect if 142* is paternally inherited (Pant, Mariano et al. 2003; Kurukuti, Tiwari et al. 2006). On the paternal chromosome, the ICR is methylated and CTCF does not bind. In mouse cells the ICR now interacts with the DMR2 site located downstream of *IGF2* and allows the enhancer to interact with the *IGF2* promoter (Murrell, Heeson et al. 2004; Kurukuti, Tiwari et al. 2006) whereas in human cells the ICR cannot interact with the CTCF DS (Nativio, Wendt et al. 2009). These results suggest that imprinting of the *H19/IGF2* locus is dependent on a specific three-dimensional organization of the region mediated by interactions between CTCF sites. Other studies suggest that CTCF-mediated interactions are not limited to the locus but also extend to CTCF sites located in other chromosomes. Using 3C or 4C, the *H19* ICR on chromosome 7 has been shown to interact with genes on different chromosomes, such as *WSB1/NF1* on chromosome 11, *ABCG2* on chromosome 6 and *OSBPL1a* on chromosome 18 (Ling, Li et al. 2006; Zhao, Tavoosidana et al. 2006). The majority of these inter-chromosomal interactions take place primarily with the maternally-inherited *H19* ICR. Knockdown of CTCF or mutation of CTCF sites on the *H19* ICR in the maternal chromosome abrogates these associations, suggesting that these interactions are CTCF dependent (Ling, Li et al. 2006; Zhao, Tavoosidana et al. 2006). Interestingly, imprinted loci are strongly overrepresented among the regions involved in inter-chromosome interactions with the *H19* ICR. The clustering of these imprinting genes by inter-chromosome interactions, which is termed 'imprinting interactome', may facilitate the regulation of these genes in *trans*. When the CTCF site is mutated in the maternal *H19* ICR, the interactions are

abrogated and the transcription of these imprinted genes is also altered; the paternal expression of the *WSB1/NF1* gene is reduced by 50-fold, expression of *Impact* is reduced by 50% and transcription of *OSBPL1a* is increased by 2.5-fold (Ling, Li et al. 2006; Zhao, Tavoosidana et al. 2006). In addition, the asynchronous replication timing of these genes in spermatogonia is switched from late to early (Sandhu, Shi et al. 2009). Therefore, CTCF mediated inter-chromosomal interactions are important for the nonallelic regulation of the epigenetic status of multiple genes in *trans*. One important advantage for such regulation is that the interactions can be reprogrammed during germline development when the epigenetic states of imprinted domains are reprogrammed. In the testis, when the maternal allele is turned to the paternal mode, the interactome can be observed in spermatogonia, but not in spermatocytes and round spermatids where the reprogramming is complete (Sandhu, Shi et al. 2009).

Another interesting example of CTCF-mediated inter-chromosome interactions takes place during X chromosome pairing, which is important for X inactivation. To achieve the mutually exclusive designation of active X (Xa) and inactive X (Xi), it is necessary for the two X chromosomes to communicate in *trans* through homologous pairing. Pairing depends on a 15-kb region within the *Tsix* and *Xite* loci. ChIP experiments show that CTCF binds both *Tsix* and *Xite* elements in female embryonic stem cells. CTCF knockdown reduces the frequency of X-X pairing to background levels in wild type embryonic stem cells and embryoid bodies (Xu, Donohoe et al. 2007). It is not clear whether CTCF mediates the pairing directly or by recruiting other factors; Oct4, which can associate with CTCF, is also needed for this pairing (Donohoe, Silva et al. 2009).

It is possible that CTCF-mediated interactions have a broader role than just regulation of transcription. One interesting example of the diverse roles that CTCF may play in nuclear biology is that of V(D)J recombination, a process that is highly regulated during B cell development. IgH rearrangement in pro-B cells begins with D_H to J_H rearrangement followed by rearrangement of a V_H gene segment to D_HJ_H. The complexity of the mechanisms controlling V_H to D_HJ_H rearrangement is in part due to the organization of the genes involved in the process. In mice, there are more than 100 V_H genes spanning a 2.5Mb region; the 96 V genes cover 3.1 Mb whereas the J genes occupy a 2 kb region. How do all the V genes access the small J cluster in the *Igh* and *Ig* loci? It has been recently shown that CTCF is present across the *Igh*, *Ig*, and *Ig* loci and plays a critical role in bringing these different loci together in the nucleus and that the cohesin subunit Rad21, which interacts with CTCF in a cell lineage-specific manner, plays a regulatory role in the process (Degner, Wong et al. 2009). Although CTCF does not display obvious cell lineage-specific binding as does Rad21, it is indispensable for the cell-lineage specific contraction of the *Igh* locus. Preliminary results for 3D-FISH indicate that *Igh* locus contraction is decreased in pro-B cells treated with CTCF shRNA (Degner-Leisso and Feeney 2010).

Taken together, these results suggest that CTCF can mediate long range intra- and inter-chromosome interactions at different loci throughout the genome either by interactions with other CTCF sites or association with other factors. The result of these interactions is a specific three-dimensional arrangement of the chromatin that can have different effects on chromosome biology, depending on the nature of the sequences brought together by CTCF.

MECHANISMS OF REGULATION OF INSULATOR FUNCTION

If insulators mediate inter- or intra-chromosomal interactions that result in the formation of chromatin loops, which in turn may be attached to the nuclear matrix, it is possible that the resulting structures determine a particular pattern of nuclear organization that may be important for gene expression. For example, it is possible that as cells differentiate,

insulator-mediated changes in nuclear organization precede or accompany cell differentiation and may be crucial in the establishment and/or maintenance of specific patterns of gene expression. If this is the case, cells must possess mechanisms to regulate insulator activity in order to establish distinct nuclear architectures that are cell fate-specific.

Evidence for the existence of mechanisms to control insulator function comes in part from genome-wide studies of insulator protein localization in *Drosophila* cell lines of different tissue origin. Studies in Kc cells, which have a neural origin, indicate that there are 3747 Su(Hw), 2266 dCTCF, 2995 BEAF and 5272 CP190 sites where these proteins are present in the genome. Of these, 47% of Su(Hw), 62% of dCTCF and 71% of BEAF sites colocalize with CP190 sites (Bushey, Ramos et al. 2009). Since CP190 is required for insulator function, this observation suggests that cells may control the activity of these various insulators by regulating the recruitment of CP190. In addition, comparison of the genomic location of different insulator proteins in Kc and Mbn2 cells (a hematopoietic cell line), reveals that while many sites are constant, a fraction of the localization sites for each of the four insulator proteins is different between the two lines. For example, 18% of Su(Hw) sites in Kc cells and 5% of Su(Hw) sites in Mbn2 cells are cell type-specific. This is also the case for dCTCF, for which 18% of sites in Kc cells and 37% in Mbn2 cells are cell-unique, whereas the number of cell type-specific BEAF sites is 11% in Kc cells and 11% in Mbn2 cells. In the case of CP190, which is found at all three insulator subclasses, 17% of sites present in Kc cells and 14% in Mbn2 cells were found to be cell type-specific (Bushey, Ramos et al. 2009). These results suggest that cells may regulate insulator activity by controlling the recruitment of the DNA binding proteins to their target sites in the genome in addition to controlling the recruitment of CP190.

Several proteins have been characterized in *Drosophila* that may play a role in regulating insulator function. dTopors, in addition to serving as an attachment point for insulators to the nuclear lamina, it has E3 ubiquitin ligase activity. This activity is required for proper insulator function. Its substrate has not been clearly identified but Su(Hw) is a likely candidate, since over-expression of dTopors enzymatic activity reverses the effect of *mod(mdg4)* mutations on the ability of Su(Hw) to interact with chromatin (Capelson and Corces 2005). In addition, modification of Mod(mdg4)2.2 and CP190 by sumoylation inhibits insulator function. Disruption of the SUMO conjugation pathway improves the enhancer-blocking function of a partially active insulator, indicating that SUMO modification acts to negatively regulate the activity of the *gypsy* insulator. Sumoylation does not affect the ability of CP190 or Mod(mdg4)2.2 to bind chromatin (Capelson and Corces 2006). Interestingly, dTopors inhibits sumoylation of Mod(mdg4)2.2 and CP190. Therefore, this protein may have a double effect on insulator function by ubiquitinating some insulator components and inhibiting the sumoylation of others. A second candidate protein with a possible role in regulating insulator function is the Rm62 RNA helicase. Insulator activity decreases in the presence of mutations in components of the RNAi machinery; insulator function is restored by mutations in Rm62. These observations have led to a model suggesting that insulator bodies contain RNA whose synthesis requires RNAi proteins. Rm62 may interact with this RNA to decrease insulator function (Lei and Corces 2006).

dTopors and Rm62 have been only shown to affect the function of the *gypsy*/Su(Hw) insulator but their potential role in regulating the activity of other insulator subclasses has not been tested. There are currently no characterized mechanisms that control the activity of the dCTCF insulator. O-glycosylation of BEAF can be detected in *Drosophila* embryonic cells in a domain of the protein that is required for association with the nuclear matrix; however, it is not clear whether glycosylation is required for *scs'* insulator function (Pathak, Rangaraj et al. 2007). A second possible candidate to regulate the BEAF insulator is the DREF protein. DREF has been characterized as a transcription factor that shares binding

sites with BEAF 32. It is possible that DREF regulates BEAF binding through competition for the same DNA sequences (Hart, Cuvier et al. 1999).

In vertebrates, the CTCF insulator can also be regulated at different levels. As in *Drosophila*, one strategy is to control the interaction between CTCF and its DNA target sequence. The best understood mechanism to control CTCF occupancy is through DNA methylation on CpG dinucleotides within and around the CTCF binding site. This has been well studied in the *H19/IGF2* locus, where CTCF only binds to the unmethylated maternal allele but does not bind to the methylated paternal allele (Bell and Felsenfeld 2000; Hark, Schoenherr et al. 2000). CTCF binding to methylation-free sites is also documented for other genes, including both imprinted and non-imprinted loci, like *RASGRF1*, *KvDMR1*, *GRB10*, *INK/ARF* and *DM1* (Filippova, Thienes et al. 2001; Hikichi, Kohda et al. 2003; Yoon, Herman et al. 2005; Fitzpatrick, Pugacheva et al. 2007; Rodriguez, Borgel et al. 2010). Therefore, regulation of CTCF binding via DNA methylation may be a general strategy to control the function of this insulator. In addition, CTCF function can be regulated by post translation modification. CTCF can be phosphorylated at the C-terminus by the protein kinase CK2. Mutation of the phosphorylated residue in CTCF does not alter its nuclear localization or DNA binding *in vitro*, but enhances repression of the *C-MYC* promoters. A phospho-mimetic mutant behaves in the opposite fashion (Klenova, Chernukhin et al. 2001; El-Kady and Klenova 2005). Thus, phosphorylation may affect the interaction of CTCF with other proteins and modulate its activity.

CTCF can also be covalently modified by poly(ADP-ribosyl)ation. CTCF interacts with Poly(ADP-ribose) polymerase (PARP)-1 both *in vivo* and *in vitro* (Guastafierro, Cecchinelli et al. 2008). Both proteins colocalize at the *IGF2/H19* ICR and overlap genome wide at more than 140 CTCF target sites in the mouse genome, including both imprinted and non-imprinted loci (Yu, Ginjala et al. 2004). Poly(ADP-ribosyl)ation does not affect the ability of CTCF to bind DNA but it is required for its insulator function. For example, mutation of the poly(ADP-ribosyl)ation site in CTCF or treatment with the PARP inhibitor ABA, can compromise imprinting of the *H19/IGF2* locus, resulting in biallelic expression of *IGF2*, but CTCF binding to the *H19/IGF2* ICR is not affected (Yu, Ginjala et al. 2004; Farrar, Rai et al. 2010). In addition to Poly(ADP-ribosyl)ation of CTCF, PARP-1 can also modify itself. The modified PARP-1 can inhibit DNMT1 activity and control DNA methylation patterns (Guastafierro, Cecchinelli et al. 2008). This may in turn control the binding of CTCF. Thus poly(ADP-ribosyl)ation may affect the function of CTCF by modulating its interaction with DNA and with other proteins.

Another mechanism to regulate insulator function that is shared by *Drosophila* and vertebrates is the use of homologous proteins that can compete for DNA binding. Like DREF and BEAF in *Drosophila*, there is also a homologue of CTCF in vertebrates named CTCFL or BORIS. These two proteins share homology of the central 12 zinc finger DNA binding domain and they can recognize the same DNA binding sequences (Loukinov, Pugacheva et al. 2002). BORIS is normally present only in the testis during germ-line development where it can stimulate the histone-methyltransferase activity of PRMT7 and contribute to the efficient DNA methylation of BORIS/CTCF sites in male germ line (Jelinic, Stehle et al. 2006). Therefore, BORIS can play a role in writing the epigenetic marks that will be read by CTCF. Usually, BORIS is expressed in a mutually exclusive manner with CTCF and only influences CTCF function by its effects in the germ line (Loukinov, Pugacheva et al. 2002). However, BORIS can be aberrantly expressed under particular conditions; for example, reduction of CTCF in normally BORIS-negative human fibroblasts results in derepression of BORIS expression (Renaud, Pugacheva et al. 2007). Abnormally expressed BORIS may affect the activity of CTCF by various mechanisms. First, BORIS can compete with CTCF for binding sites and lead to alterations in gene

expression; although BORIS shares the zinc finger domain with CTCF, the two proteins are different in the N and C terminal regions that constitute approximately two thirds of the full length amino acid sequences of these proteins and could recruit different functional partners. For example, the transcriptional factor Sp1 can interact with BORIS but not CTCF (Hong, Kang et al. 2005; Kang, Hong et al. 2007). Second, the two proteins differ in their ability to bind methylated DNA. In contrast to CTCF, the binding of BORIS to its target site is methylation-independent *in vitro* and methylation-preferential *in vivo* at the *H19/IGF2* ICR (Nguyen, Cui et al. 2008). As a consequence, BORIS could bind to methylated CTCF/BORIS sites, where CTCF is not normally present, and affect the expression of adjacent genes. Third, aberrantly expressed BORIS together with ubiquitously present PRMT7 might change the methylation status of previously non-methylated sites and evict CTCF. The binding of BORIS and the different transcription factors recruited to genes harbouring the CTCF/BORIS sites could lead to totally different expression patterns from those observed when only CTCF is present in the cell (Hong, Kang et al. 2005). Not surprisingly then, aberrant expression of BORIS can lead to disease and cancer as we discuss below.

Protein partners of CTCF may be also indispensable for the establishment and/or maintenance of chromosome interaction mediated by this protein and it is therefore plausible that regulation of these factors could change CTCF activity. Knockdown of CHD8 has been shown to impair imprinting of *IGF2* (Ishihara, Oshimura et al. 2006). Depletion of cohesin components without affecting CTCF expression can also lead to disruption of chromatin interactions and changes in the expression of genes under CTCF control (Hadjur, Williams et al. 2009; Mishiro, Ishihara et al. 2009; Nativio, Wendt et al. 2009; Hou, Dale et al. 2010). In the *Igh* locus, although the binding of CTCF is not altered during B cell differentiation, the recruitment of cohesin is cell type specific and appears to be responsible for the hypothesised role of CTCF in V(D)J recombination (Degner, Wong et al. 2009). It is therefore possible that control of the expression or post translation modification of CTCF partners will regulate the different cellular activities of CTCF. The complexity of mechanisms by which CTCF activity can be regulated is showcased by a recent report showing that RNA helicase p68 (DDX5) and its associated noncoding RNA, steroid receptor RNA activator (SRA), bind to CTCF, and are both essential *in vivo* for insulator function at the *H19/IGF2* ICR. Contrary to the inhibitory role of the RNA helicase Rm62 on insulator function in *Drosophila*, p68 is needed to positively regulate insulator function in vertebrate cells. Knockdown of p68 decreases the insulator activity of the *H19/IGF2* ICR and increases the expression of *IGF2* as previously observed in CTCF knockdown cells. In *Drosophila* cells, Rm62 interacts with CP190, while in vertebrates p68 is required for the recruitment of cohesin (Yao, Brick et al. 2010). In spite of the opposite effect of the two helicases on insulator function, the conservation of this mechanism between insects and vertebrates is striking.

INSULATORS, CANCER AND DISEASE

CTCF affects the expression of genes such as *hTERT* or *C-MYC* that regulate cell cycle processes that are important for cell growth, differentiation and apoptosis. An appropriate balance between these processes is important for normal development, whereas an imbalance can lead to tumor development. *hTERT* is the human telomerase reverse transcriptase, a catalytic subunit and limiting factor for telomerase activity, which is important during the cell cycle. CTCF binds to sequences of the *hTERT* gene located in the first two exons and represses its expression. CTCF only binds to *hTERT* in cells where the gene is not transcribed but not in telomerase-expressing cells. Knockdown of CTCF derepresses *hTERT* gene expression in normal telomerase-negative cells (Renaud, Loukinov et al. 2005). *hTERT* is a general cancer susceptibility locus, and its abnormal expression may disturb appropriate control of the cell cycle and induce aberrant cell growth and (Johnatty,

Beesley et al. 2010). CTCF also binds constitutively to regulatory elements close to the *C-MYC* P2 promoter (Gombert and Krumm 2009). Binding of CTCF can affect the expression of *C-MYC*, although the molecular mechanism underlying this effect is not well understood. Deletion of the two CTCF sites results in reduced expression of the *C-MYC* gene, while CTCF knockdown leads to an increase in C-MYC protein levels (Torrano, Chernukhin et al. 2005; Gombert and Krumm 2009). The C-MYC protein is important for cells during the transition from proliferation to differentiation. Ectopic expression of CTCF in K562 cells leads to growth retardation and promotion of differentiation associated with increased *C-MYC* expression, whereas CTCF knockdown significantly inhibits differentiation with reduced *C-MYC* transcription (Torrano, Chernukhin et al. 2005). Thus, the loss of CTCF function may disturb the balance between cell growth, differentiation and apoptosis due to its effects on the expression of cell growth regulators. This observation has been extended to other genes whose expression depends on CTCF such as the tumor repressor retinoblastoma (Rb) gene. Mutations and deletions of the Rb gene have been associated with a number of inherited malignancies. CTCF binds *in vitro* and *in vivo* to the human Rb promoter and this binding is required for Rb expression. When the CTCF binding site is removed or mutated, reporter gene expression decreases (De La Rosa-Velázquez, Rincón-Arango et al. 2007). Some other genes encoding regulators of the cell cycle, such as *p19(ARF)*, *p16(INK4a)*, *PLK*, *BRCA1*, *p53*, *p27*, and *IGF2*, are also growth suppressors frequently silenced in cancer whose expression is also controlled by CTCF (Filippova 2007). For example, *BRCA1* is expressed in normal cells but it is silenced in some cancer cells. CTCF binds at the promoter region of *BRCA1* only in expressing cells but not bind tumor cells in which *BRCA1* is silenced (Darci, Debora et al. 2004; Xu, Huo et al. 2010). The *INK4B-ARF-INK4A* (*INK/ARF*) locus contains three tumor suppressor genes that are kept silenced by DNA methylation in different types of cancer. The *p16(INK4a)* tumor suppressor gene is a frequent target of epigenetic inactivation in cancers such as breast, lung, colorectal and multiple myeloma. CTCF bind upstream of the *p16(INK4a)* promoter and the absence of binding is associated with silencing of *p16(INK4a)* expression in breast cancer and multiple myeloma cells. Moreover, ablation of CTCF protein function from *p16(INK4a)*-expressing cells by shRNA results in epigenetic changes in the *p16(INK4a)* promoter and loss of transcription (Witcher and Emerson 2009; Rodríguez, Borgel et al. 2010). Conditional over-expression of CTCF in B cells enhances expression of *p27*, *p21*, *p53*, and *p19(ARF)*, followed by inhibition of cell growth and induction of apoptosis, while knockdown of CTCF results in inhibition of these genes (Qi, Martensson et al. 2003). These results suggest that loss of CTCF function can lead to the silencing of growth suppressor genes and contribute to carcinogenesis. Whether these effects are due to changes in the three-dimensional organization of the DNA or more local effects on transcription is unclear at this time.

CTCF also plays a role in repeat-associated diseases. The affected genes in myotonic dystrophy (DM1), spinocerebellar ataxia 7 (SCA7), SCA2, dentatorubral-pallidoluysian atrophy (DRPLA) and Huntington's (HD) contain CTCF sites on one or both sides of the CTG/CAG repeat-containing region (Filippova, Thienes et al. 2001). It has been reported that CTCF can control noncoding transcription of the repeats. For example, at DM1, the two CTCF sites flanking the CTG repeat can function as an insulator to restrict the antisense transcription and constrain the heterochromatin state to the repeats without affecting the nearby DMPK and SIX5 genes, whose silencing contributes to myotonic myopathy, cataracts and cardiac-conduction defects in the disease (Filippova, Thienes et al. 2001; Cho, Thienes et al. 2005). In contrast, cells from affected individuals show a loss of CTCF binding, which is associated with expansion of the repeats, spread of heterochromatin, and regional CpG methylation (Cho, Thienes et al. 2005). Therefore, CTCF can *cis*-regulate the stability of the trinucleotide repeat. In transgenic mice carrying a SCA7 genomic fragment with a CTCF binding site, mutation or methylation of the DNA sequence at the binding site can promote repeat expansion (Libby, Hagerman et al. 2008). However, the mechanism by

which CTCF affects the stability of the repeats is not clear. A recent report suggests that CTCF can contribute to repeat stability through regulation of DNA replication. CTCF sites located between a replication origin and the repeats can slow or pause the progression of the replication machinery and enable safe passage of the replication fork. Without CTCF, the replication fork proceed in a manner that allows for slippage, hairpin formation, fork reversal and other replication errors (Cleary, Tome et al. 2010). Thus, loss of CTCF flanking the repeats may affect repeat instability. Aberrant methylation of CTCF binding sites may lead to eviction of the protein, repeat expansion and abnormal transcription of the locus.

CTCF can regulate the expression and epigenetic features of imprinted genes both in *cis* and in *trans* through intra- or inter-chromosomal interactions as discussed above. Almost all imprinted genes identified to date can be classified as regulators of embryonic growth, placental growth or adult metabolism (Jelinic and Shaw 2007). It is then not surprising that loss of imprinting can lead to various human cancers and disease. In fact, the earliest and most common alteration observed in human cancers is loss of imprinting (LOI), which has been documented in 100% of chronic myeloid leukaemia, 80% of ovarian tumors, 70% of Wilm's tumors, 66% of colorectal cancer, 56% of Barrett's oesophagus, 50% of renal-cell carcinomas, 50% of oesophageal cancer, 47–85% of lung adenocarcinoma, and 30% of meningiomas. Many human diseases are also associated with altered expression of imprinted genes, including Beckwith-Wiedemann (BWS), Prader-Willi, Angelman, Silver-Russell, immunodeficiency syndrome (ICF), Rett, Albright hereditary osteodystrophy, and hydatidiform mole (Jelinic and Shaw 2007). Although the role of CTCF on imprinting has only been studied in detail in the *H19/IGF2* locus, similar CTCF-associated mechanisms appear to also play a role in imprinting at the *RASGRF1*, *DLK1/GTL2*, *WSB1/NF1* and *KvDMR* loci (Yoon, Herman et al. 2005; Ling, Li et al. 2006; Fitzpatrick, Pugacheva et al. 2007). This suggests a widespread function of CTCF in the regulation of genomic imprinting and, as a consequence, the development of imprinting-associated human cancer and other diseases.

Alterations of CTCF function that lead to disease development can have both genetic and epigenetic origins. The CTCF gene maps to the cancer-associated human chromosome locus 16q22.1, which is the smallest overlap region of a variety of deletions found in breast, prostate, ovarian and Wilm's tumors. Since CTCF null mutations are lethal, most CTCF mutations involved in human disease appear to cluster in the zinc finger domain of the protein. Instead of causing a complete loss of function, mutations in this domain alter the binding ability of CTCF. For example, one of the mutations identified abrogates CTCF binding to a subset of target sites in certain genes involved in the regulation of cell proliferation (*C-MYC*, *ARF*, *PIMI*, *PLK*, and *IGF2*) but does not alter binding to other regions of the genome, including the β -globin insulator, the lysozyme silencer, or the APP promoter (Filippova, Qi et al. 2002). The selective loss of CTCF function caused by mutations in the zinc finger region may be tolerated because it does not affect cell viability, but results instead in transformation to a malignant phenotype. Since only selective changes of CTCF function can be tolerated in cells, it is not surprising that CTCF mutations are infrequently discovered in cancer or disease states. Instead, it is possible that mutations leading to changes of CTCF function could affect regulation of CTCF function rather than its structure. This alterations could be epigenetic, gene specific and relatively tolerable. Most CTCF-mediated cancers may occur through aberrant methylation of CTCF binding sites, which results in loss of CTCF binding to the DNA. For example, abnormal methylation patterns of CTCF targets involved in *H19/IGF2* imprinting control have been shown to be associated with colorectal, Wilm's, and bladder tumors (Klenova, Morse et al. 2002). The methylation pattern at the *KvDMR* locus was found to be altered in a child with Beckwith-Wiedemann syndrome (Fitzpatrick, Pugacheva et al. 2007). This alteration of methylation at CTCF sites has also been observed at non-imprinted genes. For example, in mice lung

tumors, CpG sites harbouring CTCF binding sequences are hypermethylated at the *INK/ARF* locus, resulting in the absence of CTCF binding and reduced expression of the tumor suppressor genes at this locus (Tam, Devereux et al. 2003; Rodriguez, Borgel et al. 2010). In some human tumors, the promoter of the *BRCA1* gene is methylated, CTCF is evicted and *BRCA1* is not expressed (Darci, Debora et al. 2004; Xu, Huo et al. 2010).

In addition to DNA methylation, other regulatory mechanisms of CTCF function can be altered and lead to cancer and disease. For example, testicular and ovarian tumor cells exhibit the same methylation profile as normal cells, but the BORIS protein, which recognizes the same DNA binding sites as CTCF, is expressed abnormally. Both CTCF and BORIS bind to DNA in cancer cells but BORIS inhibits the function of CTCF and leads to expression of *hTERT* (Renaud, Loukinov et al. 2010). Telomerase activity is not detectable in most somatic cells of adult humans but is found in highly proliferative cells, such as germ cells and stem cells, and 85–95% of cancers. The expression of BORIS in normal cells is sufficient to allow *hTERT* transcription and to extend their lifespan *in vitro* (Renaud, Loukinov et al. 2010). Reciprocal binding of CTCF and BORIS has also been observed at the *NY-ESO-1* promoter and leads to derepression of this cancer-testis gene in lung tumors (Hong, Kang et al. 2005). BORIS is also aberrantly expressed in 71 % (41 of 58 cases) of breast tumors. High levels of BORIS correlate with high levels of progesterone receptor (PR) and oestrogen receptor (ER). The link between BORIS and PR/ER was further confirmed by the ability of BORIS to activate the promoters of the *PR* and *ER* genes in reporter gene assays (D'Arcy, Pore et al. 2008). Defects in other regulatory mechanisms that alter CTCF function can also lead to cancer. For example, alterations in the poly(ADP-ribosylation) pathway results in the absence of CTCF PARlation and reduction of *p16(INK4a)* and *RASSF1A* expression (Witcher and Emerson 2009).

Understanding of the different mechanism that alter CTCF function has opened new possibilities in the design of treatments for cancer or other diseases. For example, Trichostatin A (TSA), an inhibitor of histone deacetylase activity, is a well-known antitumor agent that effectively and selectively induces arrest of tumor growth and apoptosis. *hTERT* appears to be one of the primary targets for TSA-induced apoptosis in cancer cells. TSA induces demethylation of CpGs present in the binding site of CTCF on the *hTERT* promoter, leading to repression of *hTERT* (Choi, Min et al. 2010). CTCF can contribute to the regulation of a variety of genes whose proper expression is required for normal cell differentiation. Both genetic and epigenetic changes of CTCF function can lead to the mis-expression of these genes, resulting in the development of a malignant phenotype. Understanding the mechanisms by which CTCF function can be regulated may result in the development of novel methods to treat cancer and other diseases.

FUTURE QUESTIONS

Chromatin insulators are important regulatory sequences present in the genome of most eukaryotes. Although they are defined experimentally based on their ability to affect enhancer-promoter interactions and interfere with the spreading of repressive signals from heterochromatin, their role appears to be more general. Intra- and inter-chromosomal interactions mediated by insulator proteins may establish a web of contacts between individual insulator sites that give rise to specific patterns of nuclear organization. Insulator-mediated nuclear structures may be regulatable by controlling the interaction between insulator DNA binding proteins and their cognate sequences. In addition, recruitment of insulator components involved in mediating inter-insulator interactions may represent a second level of controlling the function of these elements. These two levels of regulation may be the result of specific covalent modifications of insulator proteins. The specific outcome of inter-insulator interactions is a consequence of the location of the particular

insulator sequences with respect to specific genome features, and interference with enhancer-promoter interactions may be just one of these outcomes. It is unclear at this time whether the three-dimensional pattern of nuclear organization created by the insulator interactome is different in various cell types and whether this organization carries epigenetic information. Most studies to date have concentrated on the analysis of alterations of insulator function at the local level. These studies suggest that aberrant insulator function can lead to alterations in gene expression and, when the affected genes are involved in cell growth processes, to cancer. Nevertheless, it is possible that changes in insulator function cause more global effects on transcription. Understanding the role of nuclear organization in gene expression and cell differentiation remains the main issue for future investigation.

Acknowledgments

We would like to thank Chunhui Hou for inspirational discussions. Work in the authors' laboratory is funded by National Institutes of Health Award GM35463.

References

- Barski A, Cuddapah S, et al. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*. 2007; 129(4):823–837. [PubMed: 17512414]
- Bartkuhn M, Straub T, et al. Active promoters and insulators are marked by the centrosomal protein 190. *Embo J*. 2009; 28(7):877–88. [PubMed: 19229299]
- Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*. 2000; 405(6785):482–485. [PubMed: 10839546]
- Belozero V, Majumder P, et al. A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of *Drosophila*. *Embo J*. 2003; 22(12):3113–21. [PubMed: 12805225]
- Blanton J, Gaszner M, et al. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev*. 2003; 17(5):664–75. [PubMed: 12629048]
- Buchner K, Roth P, et al. Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in *Drosophila*. *Genetics*. 2000; 155(1):141–57. [PubMed: 10790390]
- Bushey AM, Dorman ER, et al. Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol Cell*. 2008; 32(1):1–9. [PubMed: 18851828]
- Bushey AM, Ramos E, et al. Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions. *Genes Dev*. 2009; 23(11):1338–50. [PubMed: 19443682]
- Byrd K V, Corces G. Visualization of chromatin domains created by the gypsy insulator of *Drosophila*. *J Cell Biol*. 2003; 162(4):565–74. [PubMed: 12925706]
- Capelson M V, Corces G. The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol Cell*. 2005; 20(1):105–16. [PubMed: 16209949]
- Capelson M, Corces VG. SUMO conjugation attenuates the activity of the gypsy chromatin insulator. *EMBO J*. 2006; 25(9):1906–14. [PubMed: 16628226]
- Chen X, Xu H, et al. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*. 2008; 133(6):1106–1117. [PubMed: 18555785]
- Chernukhin I, Shamsuddin S, et al. CTCF Interacts with and Recruits the Largest Subunit of RNA Polymerase II to CTCF Target Sites Genome-Wide. *Mol Cell Biol*. 2007; 27(5):1631–1648. [PubMed: 17210645]
- Chernukhin IV, Shamsuddin S, et al. Physical and Functional Interaction between Two Pluripotent Proteins, the Y-box DNA/RNA-binding Factor, YB-1, and the Multivalent Zinc Finger Factor, CTCF. *Journal of Biological Chemistry*. 2000; 275(38):29915–29921. [PubMed: 10906122]
- Cho DH, Thienes CP, et al. Antisense Transcription and Heterochromatin at the DM1 CTG Repeats Are Constrained by CTCF. *Molecular Cell*. 2005; 20(3):483–489. [PubMed: 16285929]
- Choi JH, Min NY, et al. TSA-induced DNMT1 down-regulation represses hTERT expression via recruiting CTCF into demethylated core promoter region of hTERT in HCT116. *Biochemical and Biophysical Research Communications*. 2010; 391(1):449–454. [PubMed: 19914212]

- Cleary JD, Tome S, et al. Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat Struct Mol Biol.* 2010; 17(99):1079–1087. [PubMed: 20711191]
- Cuddapah S, Jothi R, et al. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Research.* 2009; 19(1):24–32. [PubMed: 19056695]
- D'Arcy V, Pore N, et al. BORIS, a paralogue of the transcription factor, CTCF, is aberrantly expressed in breast tumours. *Br J Cancer.* 2008; 98(3):571–579. [PubMed: 18195709]
- Darci TB, Debora NMD, et al. DNA binding sites for putative methylation boundaries in the unmethylated region of the <I>BRCA1</I> promoter. *International Journal of Cancer.* 2004; 111(5):669–678.
- De La Rosa-Velázquez IA, Rincón-Arango H, et al. Epigenetic Regulation of the Human Retinoblastoma Tumor Suppressor Gene Promoter by CTCF. *Cancer Research.* 2007; 67(6):2577–2585. [PubMed: 17363576]
- Defossez PA, Kelly KF, et al. The Human Enhancer Blocker CTC-binding Factor Interacts with the Transcription Factor Kaiso. *Journal of Biological Chemistry.* 2005; 280(52):43017–43023. [PubMed: 16230345]
- Degner-Leisso SC, Feeney AJ. Epigenetic and 3-dimensional regulation of V(D)J rearrangement of immunoglobulin genes. *Seminars in Immunology.* 2010 In Press, Corrected Proof.
- Degner SC, Wong TP, et al. Cutting Edge: Developmental Stage-Specific Recruitment of Cohesin to CTCF Sites throughout Immunoglobulin Loci during B Lymphocyte Development. *J Immunol.* 2009; 182(1):44–48. [PubMed: 19109133]
- Dhillon N, Raab J, et al. DNA polymerase epsilon, acetylases and remodellers cooperate to form a specialized chromatin structure at a tRNA insulator. *Embo J.* 2009; 28(17):2583–600. [PubMed: 19629037]
- Donohoe ME, Silva SS, et al. The pluripotency factor Oct4 interacts with Ctf and also controls X-chromosome pairing and counting. *Nature.* 2009; 460(7251):128–132. [PubMed: 19536159]
- Donohoe ME, Zhang LF, et al. Identification of a Ctf Cofactor, Yy1, for the X Chromosome Binary Switch. *Molecular Cell.* 2007; 25(1):43–56. [PubMed: 17218270]
- El-Kady A, Klenova E. Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. *FEBS Letters.* 2005; 579(6):1424–1434. [PubMed: 15733852]
- Emberly E, Blattes R, et al. BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites. *PLoS Biol.* 2008; 6(12):2896–910. [PubMed: 19108610]
- Farrar D, Rai S, et al. Mutational Analysis of the Poly(ADP-Ribosyl)ation Sites of the Transcription Factor CTCF Provides an Insight into the Mechanism of Its Regulation by Poly(ADP-Ribosyl)ation. *Mol Cell Biol.* 2010; 30(5):1199–1216. [PubMed: 20038529]
- Farrell CM, Grinberg A, et al. A large upstream region is not necessary for gene expression or hypersensitive site formation at the mouse β -globin locus. *Proceedings of the National Academy of Sciences of the United States of America.* 2000; 97(26):14554–14559. [PubMed: 11121056]
- Ferraiuolo MA, Rousseau M, et al. The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Research.* 2010
- Filippova, GN. Genetics and Epigenetics of the Multifunctional Protein CTCF. In: Gerald, PS., editor. *Current Topics in Developmental Biology.* Vol. 80. Academic Press; 2007. p. 337-360.
- Filippova GN, Fagerlie S, et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol.* 1996; 16(6):2802–13. [PubMed: 8649389]
- Filippova GN, Qi CF, et al. Tumor-associated Zinc Finger Mutations in the CTCF Transcription Factor Selectively Alter Its DNA-binding Specificity. *Cancer Research.* 2002; 62(1):48–52. [PubMed: 11782357]
- Filippova GN, Thienes CP, et al. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat Genet.* 2001; 28(4):335–343. [PubMed: 11479593]
- Fitzpatrick GV, Pugacheva EM, et al. Allele-Specific Binding of CTCF to the Multipartite Imprinting Control Region KvDMR1. *Mol Cell Biol.* 2007; 27(7):2636–2647. [PubMed: 17242189]

- Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet.* 2006; 7(9):703–13. [PubMed: 16909129]
- Gaszner M, Vazquez J, et al. The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes Dev.* 1999; 13(16):2098–107. [PubMed: 10465787]
- Gause M, Morcillo P, et al. Insulation of enhancer-promoter communication by a gypsy transposon insert in the *Drosophila* cut gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. *Mol Cell Biol.* 2001; 21(14):4807–17. [PubMed: 11416154]
- Gerasimova TI, Byrd K, et al. A chromatin insulator determines the nuclear localization of DNA. *Mol Cell.* 2000; 6(5):1025–35. [PubMed: 11106742]
- Gerasimova TI, Gdula DA, et al. A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell.* 1995; 82(4):587–97. [PubMed: 7664338]
- Gerasimova TI, Lei EP, et al. Coordinated control of dCTCF and gypsy chromatin insulators in *Drosophila*. *Mol Cell.* 2007; 28(5):761–72. [PubMed: 18082602]
- Geyer PK V, Corces G. DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Genes & Development.* 1992; 6(10):1865–1873. [PubMed: 1327958]
- Ghosh D, Gerasimova TI, et al. Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. *Embo J.* 2001; 20(10):2518–27. [PubMed: 11350941]
- Gilbert MK, Tan YY, et al. The *Drosophila* boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. *Genetics.* 2006; 173(3):1365–75. [PubMed: 16648647]
- Gombert WM, Krumm A. Targeted Deletion of Multiple CTCF-Binding Elements in the Human C-MYC Gene Reveals a Requirement for CTCF in C-MYC Expression. *PLoS ONE.* 2009; 4(7):e6109. [PubMed: 19568426]
- Guastafierro T, Cecchinelli B, et al. CCCTC-binding Factor Activates PARP-1 Affecting DNA Methylation Machinery. *Journal of Biological Chemistry.* 2008; 283(32):21873–21880. [PubMed: 18539602]
- Guelen L, Pagie L, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature.* 2008; 453(7197):948–951. [PubMed: 18463634]
- Gyurkovics H, Gausz J, et al. A new homeotic mutation in the *Drosophila* bithorax complex removes a boundary separating two domains of regulation. *The EMBO journal.* 1990; 9(8):2579–85. [PubMed: 1973385]
- Hadjur S, Williams LM, et al. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature.* 2009; 460(7253):410–413. [PubMed: 19458616]
- Hark AT, Schoenherr CJ, et al. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature.* 2000; 405(6785):486–489. [PubMed: 10839547]
- Harrison DA, Gdula DA, et al. A leucine zipper domain of the suppressor of Hairy-wing protein mediates its repressive effect on enhancer function. *Genes Dev.* 1993; 7(10):1966–78. [PubMed: 7916729]
- Hart CM, Cuvier O, et al. Evidence for an antagonistic relationship between the boundary element-associated factor BEAF and the transcription factor DREF. *Chromosoma.* 1999; 108(6):375–83. [PubMed: 10591997]
- Heath H, Ribeiro de Almeida C, et al. CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *Embo J.* 2008; 27(21):2839–50. [PubMed: 18923423]
- Hikichi T, Kohda T, et al. Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucl Acids Res.* 2003; 31(5):1398–1406. [PubMed: 12595547]
- Holohan EE, Kwong C, et al. CTCF genomic binding sites in *Drosophila* and the organisation of the bithorax complex. *PLoS Genet.* 2007; 3(7):e112. [PubMed: 17616980]
- Hong JA, Kang Y, et al. Reciprocal Binding of CTCF and BORIS to the NY-ESO-1 Promoter Coincides with Derepression of this Cancer-Testis Gene in Lung Cancer Cells. *Cancer Res.* 2005; 65(17):7763–7774. [PubMed: 16140944]
- Hoover KK, Chien AJ, et al. Effects of Transposable Elements on the Expression of the forked Gene of *Drosophila melanogaster*. *Genetics.* 1993; 135(2):507–526. [PubMed: 8244011]

- Hou C, Dale R, et al. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proceedings of the National Academy of Sciences*. 2010; 107(8):3651–3656.
- Huang S, Li X, et al. USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Mol Cell Biol*. 2007; 27(22):7991–8002. [PubMed: 17846119]
- Ishihara K, Oshimura M, et al. CTCF-Dependent Chromatin Insulator Is Linked to Epigenetic Remodeling. *Molecular Cell*. 2006; 23(5):733–742. [PubMed: 16949368]
- Jelinic P, Shaw P. Loss of imprinting and cancer. *The Journal of Pathology*. 2007; 211(3):261–268. [PubMed: 17177177]
- Jelinic P, Stehle JC, et al. The Testis-Specific Factor CTCFL Cooperates with the Protein Methyltransferase PRMT7 in H19 Imprinting Control Region Methylation. *PLoS Biol*. 2006; 4(11):e355. [PubMed: 17048991]
- Jiang N, Emberly E, et al. Genome-wide mapping of BEAF binding sites in *Drosophila* links BEAF to transcription. *Mol Cell Biol*. 2009
- Johnatty SE, Beesley J, et al. Evaluation of Candidate Stromal Epithelial CrossTalk Genes Identifies Association between Risk of Serous Ovarian Cancer and *TERT*, a Cancer Susceptibility “Hot-Spot”. *PLoS Genet*. 2010; 6(7):e1001016. [PubMed: 20628624]
- Jothi R, Cuddapah S, et al. Genome-wide identification of in vivo protein – DNA binding sites from ChIP-Seq data. *Nucleic Acids Research*. 2008; 36(16):5221–5231. [PubMed: 18684996]
- Kang Y, Hong JA, et al. Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells. *Oncogene*. 2007; 26(30):4394–4403. [PubMed: 17260018]
- Karch FO, Galloni M, et al. Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of *Drosophila melanogaster*. *Nucleic Acids Research*. 1994; 22(15):3138–3146. [PubMed: 7915032]
- Kellum R, Schedl P. A group of scs elements function as domain boundaries in an enhancer- blocking assay. *Mol Cell Biol*. 1992; 12(5):2424–31. [PubMed: 1569958]
- Kim J, Kollhoff A, et al. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Hum Mol Genet*. 2003; 12(3): 233–245. [PubMed: 12554678]
- Kim TH, Abdullaev ZK, et al. Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell*. 2007; 128(6):1231–1245. [PubMed: 17382889]
- Klenova EM I, Chernukhin V, et al. Functional Phosphorylation Sites in the C-Terminal Region of the Multivalent Multifunctional Transcriptional Factor CTCF. *Mol Cell Biol*. 2001; 21(6):2221–2234. [PubMed: 11238955]
- Klenova EM, Morse HC, et al. The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Seminars in Cancer Biology*. 2002; 12(5):399–414. [PubMed: 12191639]
- Kurukuti S V, Tiwari K, et al. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences*. 2006; 103(28):10684–10689.
- Labrador M V, Corces G. Extensive exon reshuffling over evolutionary time coupled to trans-splicing in *Drosophila*. *Genome Res*. 2003; 13(10):2220–8. [PubMed: 14525924]
- Labrador M, Mongelard F, et al. Protein encoding by both DNA strands. *Nature*. 2001; 409:1000. [PubMed: 11234000]
- Lei EP V, Corces G. RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat Genet*. 2006; 38(8):936–41. [PubMed: 16862159]
- Libby RT, Hagerman KA, et al. CTCF cis-Regulates Trinucleotide Repeat Instability in an Epigenetic Manner: A Novel Basis for Mutational Hot Spot Determination. *PLoS Genet*. 2008; 4(11):e1000257. [PubMed: 19008940]
- Ling JQ, Li T, et al. CTCF Mediates Interchromosomal Colocalization Between Igf2/H19 and Wsb1/Nf1. *Science*. 2006; 312(5771):269–272. [PubMed: 16614224]
- Loukinov DI, Pugacheva E, et al. BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator

- protein involved in reading imprinting marks in the soma. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99(10):6806–6811. [PubMed: 12011441]
- Lutz M, Burke LJ, et al. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucl Acids Res*. 2000; 28(8):1707–1713. [PubMed: 10734189]
- Majumder P, Gomez JA, et al. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *The Journal of Experimental Medicine*. 2008; 205(4):785–798. [PubMed: 18347100]
- Melnikova L, Juge F, et al. Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in *Drosophila*. *Proc Natl Acad Sci U S A*. 2004; 101(41):14806–11. [PubMed: 15465920]
- Mishiro T, Ishihara K, et al. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J*. 2009; 28(9):1234–1245. [PubMed: 19322193]
- Mohan M, Bartkuhn M, et al. The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *Embo J*. 2007; 26(19):4203–14. [PubMed: 17805343]
- Murrell A, Heeson S, et al. Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat Genet*. 2004; 36(8):889–893. [PubMed: 15273689]
- Nativio R, Wendt KS, et al. Cohesin Is Required for Higher-Order Chromatin Conformation at the Imprinted *IGF2-H19* Locus. *PLoS Genet*. 2009; 5(11):e1000739. [PubMed: 19956766]
- Nguyen P, Cui H, et al. CTCFL/BORIS Is a Methylation-Independent DNA-Binding Protein That Preferentially Binds to the Paternal *H19* Differentially Methylated Region. *Cancer Research*. 2008; 68(14):5546–5551. [PubMed: 18632606]
- Ohlsson R, Renkawitz R, et al. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet*. 2001; 17(9):520–7. [PubMed: 11525835]
- Ottaviani A, Rival-Gervier S, et al. The *D4Z4* Macrosatellite Repeat Acts as a CTCF and A-Type Lamins-Dependent Insulator in Facio-Scapulo-Humeral Dystrophy. *PLoS Genet*. 2009; 5(2):e1000394. [PubMed: 19247430]
- Ottaviani A, Schluth-Bolard C, et al. Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. *EMBO J*. 2009; 28(16):2428–2436. [PubMed: 19644448]
- Pai CY, Lei EP, et al. The Centrosomal Protein CP190 Is a Component of the gypsy Chromatin Insulator. *Mol Cell*. 2004; 16(5):737–48. [PubMed: 15574329]
- Pant V, Mariano P, et al. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the *H19* imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes & Development*. 2003; 17(5):586–590. [PubMed: 12629040]
- Parelho V, Hadjur S, et al. Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell*. 2008; 132(3):422–433. [PubMed: 18237772]
- Pathak RU, Rangaraj N, et al. Boundary element-associated factor 32B connects chromatin domains to the nuclear matrix. *Mol Cell Biol*. 2007; 27(13):4796–806. [PubMed: 17485444]
- Phillips JE V, Corces G. CTCF: master weaver of the genome. *Cell*. 2009; 137(7):1194–211. [PubMed: 19563753]
- Qi CF, Martensson A, et al. CTCF functions as a critical regulator of cell-cycle arrest and death after ligation of the B cell receptor on immature B cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(2):633–8. [PubMed: 12524457]
- Renaud S, Loukinov D, et al. BORIS/CTCF-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells. *Nucleic Acids Research*. 2010
- Renaud S, Loukinov D, et al. CTCF binds the proximal exonic region of hTERT and inhibits its transcription. *Nucl Acids Res*. 2005; 33(21):6850–6860. [PubMed: 16326864]
- Renaud S, Pugacheva EM, et al. Expression of the CTCF-paralogous cancer-testis gene, brother of the regulator of imprinted sites (BORIS), is regulated by three alternative promoters modulated by CpG methylation and by CTCF and p53 transcription factors. *Nucl Acids Res*. 2007; 35(21):7372–7388. [PubMed: 17962299]

- Rodriguez C, Borgel J, et al. CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochemical and Biophysical Research Communications*. 2010; 392(2):129–134. [PubMed: 20051228]
- Roy S, Gilbert MK, et al. Characterization of BEAF mutations isolated by homologous recombination in *Drosophila*. *Genetics*. 2007; 176(2):801–13. [PubMed: 17435231]
- Roy S, Tan YY, et al. A genetic screen supports a broad role for the *Drosophila* insulator proteins BEAF-32A and BEAF-32B in maintaining patterns of gene expression. *Mol Genet Genomics*. 2007; 277(3):273–86. [PubMed: 17143631]
- Rubio ED, Reiss DJ, et al. CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences*. 2008; 105(24):8309–8314.
- Sandhu KS, Shi C, et al. Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes & Development*. 2009; 23(22):2598–2603. [PubMed: 19933149]
- Schweinsberg S, Hagstrom K, et al. The enhancer-blocking activity of the Fab-7 boundary from the *Drosophila* bithorax complex requires GAGA-factor-binding sites. *Genetics*. 2004; 168(3):1371–84. [PubMed: 15579691]
- Sekimata M, Pérez-Melgosa M, et al. CCCTC-Binding Factor and the Transcription Factor T-bet Orchestrate T Helper 1 Cell-Specific Structure and Function at the Interferon- γ Locus. *Immunity*. 2009; 31(4):551–564. [PubMed: 19818655]
- Splinter E, Heath H, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes & Development*. 2006; 20(17):2349–2354. [PubMed: 16951251]
- Splinter E, Heath H, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 2006; 20(17):2349–54. [PubMed: 16951251]
- Stedman W, Kang H, et al. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J*. 2008; 27(4):654–666. [PubMed: 18219272]
- Tam AS, Devereux TR, et al. Perturbations of the Ink4a/Arf gene locus in aflatoxin B1-induced mouse lung tumors. *Carcinogenesis*. 2003; 24(1):121–132. [PubMed: 12538357]
- Tolhuis B, Palstra RJ, et al. Looping and Interaction between Hypersensitive Sites in the Active beta-globin Locus. *Molecular Cell*. 2002; 10(6):1453–1465. [PubMed: 12504019]
- Torrano V, Chernukhin I, et al. CTCF Regulates Growth and Erythroid Differentiation of Human Myeloid Leukemia Cells. *Journal of Biological Chemistry*. 2005; 280(30):28152–28161. [PubMed: 15941718]
- Torrano VN, Chernukhin I, et al. CTCF Regulates Growth and Erythroid Differentiation of Human Myeloid Leukemia Cells. *Journal of Biological Chemistry*. 2005; 280(30):28152–28161. [PubMed: 15941718]
- Wallace JA, Felsenfeld G. We gather together: insulators and genome organization. *Curr Opin Genet Dev*. 2007; 17(5):400–7. [PubMed: 17913488]
- Wendt KS, Yoshida K, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*. 2008; 451(7180):796–801. [PubMed: 18235444]
- Weth O, Weth C, et al. Modular Insulators: Genome Wide Search for Composite CTCF/Thyroid Hormone Receptor Binding Sites. *PLoS ONE*. 2010; 5(4):e10119. [PubMed: 20404925]
- Witcher M, Emerson BM. Epigenetic Silencing of the p16INK4a Tumor Suppressor Is Associated with Loss of CTCF Binding and a Chromatin Boundary. *Molecular Cell*. 2009; 34(3):271–284. [PubMed: 19450526]
- Xu J, Huo D, et al. CpG island methylation affects accessibility of the proximal BRCA1 promoter to transcription factors. *Breast Cancer Research and Treatment*. 2010; 120(3):593–601. [PubMed: 19466541]
- Xu N, Donohoe ME, et al. Evidence that homologous X-chromosome pairing requires transcription and Ctf protein. *Nat Genet*. 2007; 39(11):1390–1396. [PubMed: 17952071]
- Yao H, Brick K, et al. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes & Development*. 2010
- Yoon B, Herman H, et al. Rasgrf1 Imprinting Is Regulated by a CTCF-Dependent Methylation-Sensitive Enhancer Blocker. *Mol Cell Biol*. 2005; 25(24):11184–11190. [PubMed: 16314537]

- Yu W, Ginjala V, et al. Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat Genet.* 2004; 36(10):1105–1110. [PubMed: 15361875]
- Yusufzai TM, Tagami H, et al. CTCF Tethers an Insulator to Subnuclear Sites, Suggesting Shared Insulator Mechanisms across Species. *Molecular Cell.* 2004; 13(2):291–298. [PubMed: 14759373]
- Zhao K, Hart CM, et al. Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell.* 1995; 81(6):879–89. [PubMed: 7781065]
- Zhao Z, Tavoosidana G, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet.* 2006; 38(11):1341–1347. [PubMed: 17033624]

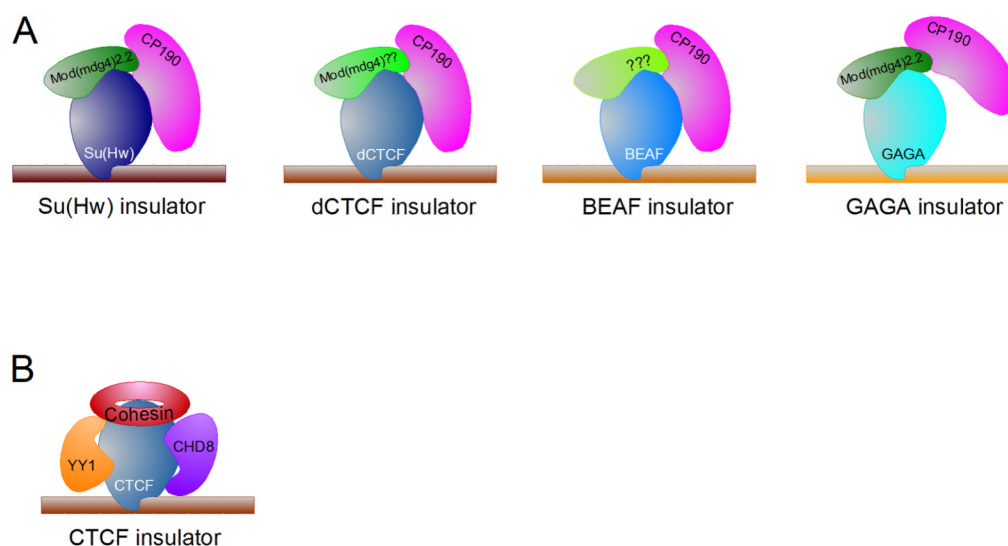
**Figure 1.**

Diagram showing the structure of different *Drosophila* and vertebrate insulators. **A.** Each *Drosophila* insulator subclass contains a different binding protein that may define the specific function of the corresponding subclass. All insulators share the common protein CP190, although the role of this protein in the function of the GAGA insulator has not been demonstrated experimentally. In addition, all subclasses may also have one Mod(mdg4) isoform. The gypsy/Su(Hw) insulator contains Mod(mdg4)2.2. The dCTCF insulator lacks this isoform but contains a different variant of Mod(mdg4) (T. Gerasimova, E. Lei and V. Corces, unpublished observations). It is not known whether the BEAF insulator contains a Mod(mdg4) variant but GAGA has been shown to interact with Mod(mdg4)2.2. **B.** Structure of the vertebrate CTCF insulator.

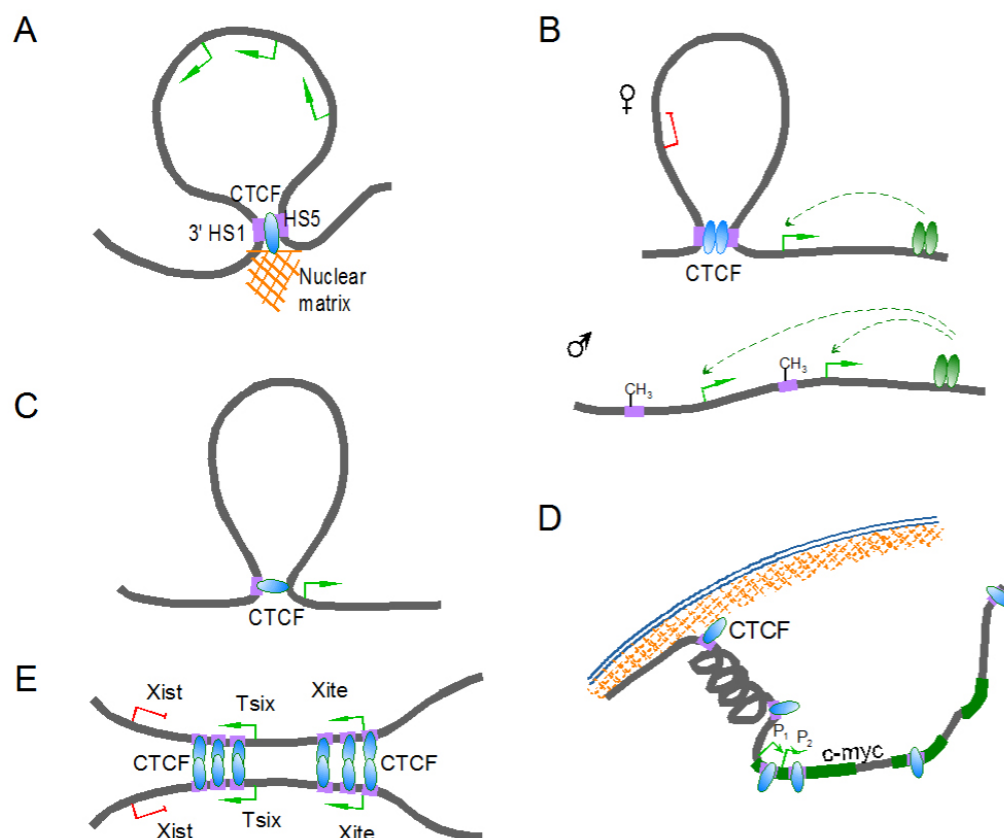
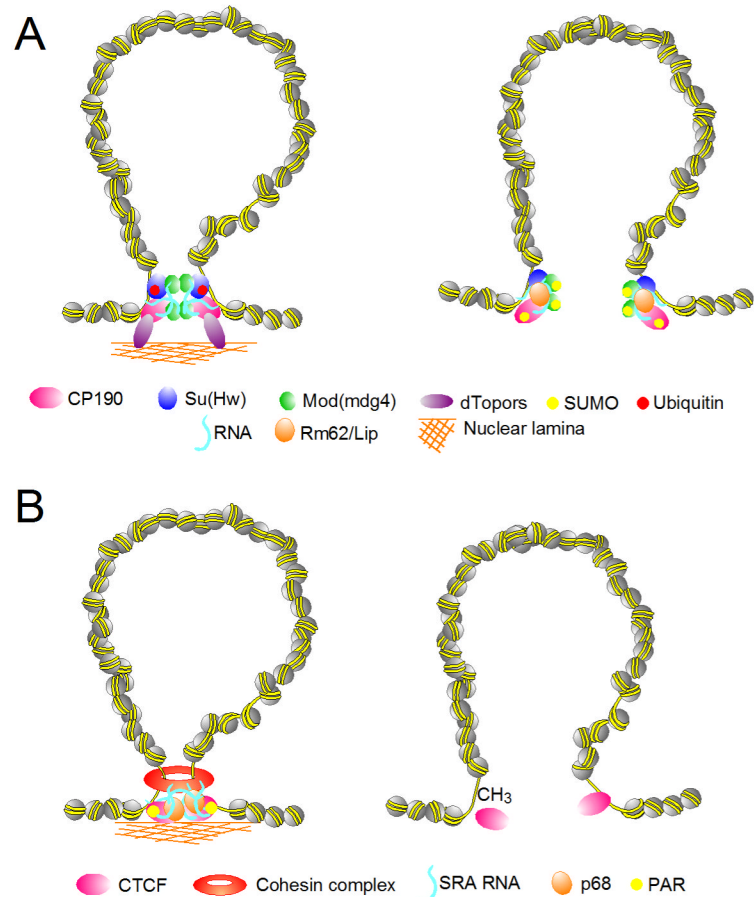


Figure 2. Mechanisms by which CTCF affects gene expression by mediating intra- and inter-chromosomal interactions. **A.** Anchoring loops via direct attachment to subnuclear structures such as the nucleolus and/or nuclear matrix. **B.** Monoallelic gene expression via allele-specific contacts between multiple imprinted regulatory elements. **C.** Transcriptional regulation via contacts between intergenic enhancers and promoter-proximal regulatory elements. **D.** Global nuclear organization via demarcation of lamina-associated domains (LADs). **E.** X chromosome inactivation or monoallelic gene expression via inter-chromosomal contacts between regulatory elements in *trans*.

**Figure 3.**

Similarities in the regulatory mechanisms of insulator function in *Drosophila* and vertebrates. **A.** Regulation of insulator activity in *Drosophila*. The left panel shows two active insulators coming together to make a functional loop and the right panel displays two inactive insulators unable to form a loop. At an active insulator site, dTopors is present, Rm62/Lip is not present, Su(Hw) is ubiquitinated, Mod(mdg4)2.2 and CP190 are not sumoylated and dTopors serves as a bridge to the nuclear lamina/matrix. At inactive insulator sites, dTopors is absent and Su(Hw) is not ubiquitinated, whereas Mod(mdg4)2.2 and CP190 are sumoylated. Rm62/Lip is present and bound to RNA. Under these conditions, the two insulator sites cannot interact to form a loop. Absence of dTopors also precludes interactions with the nuclear lamina. **B.** Regulation of insulator activity in vertebrates. The left panel shows two active CTCF insulators coming together to make a functional loop and the right panel displays two inactive CTCF insulators unable to form a loop. At an active insulator site, CTCF is present and poly(ADP-ribosyl)ated and, directly or indirectly, interacts with the nuclear matrix/lamina; in addition, the SRA RNA and p68 are present while the cohesin complex brings together the two DNA molecules forming the base of the loop. At inactive insulator sites, CTCF is either not bound because of the DNA is methylated or it is bound but not modified by poly(ADP-ribosyl)ation; cohesin, SRA RNA and/or p68 are absent. Under these conditions the two insulator sites cannot interact to form a loop.