CTCF and its protein partners: divide and rule?

Jordanka Zlatanova[1,*] and Paola Caiafa[2]

[1]Department of Molecular Biology, University of Wyoming, Laramie, WY 82071, USA
[2]Department of Cellular Biotechnology and Hematology, Second Faculty of Medicine and Surgery, University 'La Sapienza', and Pasteur Institute-Fondazione Cenci Bolognetti, 00161 Rome, Italy
*Author for correspondence (e-mail: jordanka@uwyo.edu)

Journal of Cell Science 122, 1275-1284 Published by The Company of Biologists 2009
doi:10.1242/jcs.039990

Summary

CTCF is a ubiquitous transcription factor that is involved in numerous, seemingly unrelated functions. These functions include, but are not limited to, positive or negative regulation of transcription, enhancer-blocking activities at developmentally regulated gene clusters and at imprinted loci, and X-chromosome inactivation. Here, we review recent data acquired with state-of-the-art technologies that illuminate possible mechanisms behind the diversity of CTCF functions. CTCF interacts with numerous protein partners, including cohesin, nucleophosmin, PARP1, Yy1 and RNA polymerase II. We propose that CTCF interacts with one or two different partners according to the biological context, applying the Roman principle of governance, 'divide and rule' (divide et impera).

Key words: Cohesin, Nucleophosmin, PARP1, RNA polymerase II, Yy1

Introduction

CCCTC-binding factor (CTCF) is a ubiquitously expressed 11-zinc-finger vertebrate protein that binds to thousands of sites in the genome in a sequence-specific manner and performs myriad functions. Initially, CTCF was described as a transcriptional repressor of the *Myc* gene; later studies, however, recognized its involvement in very diverse functions, including enhancer blocking, X-chromosome inactivation, gene imprinting and promoter activation or repression (Fig. 1A) (for reviews, see Ohlsson et al., 2001; Gazner and Felsenfeld, 2006; Wallace and Felsenfeld, 2007; Filippova, 2008).

How can one ubiquitous protein perform so many functions, which are often seemingly unrelated? The answer might lie in the context-dependent interactions of CTCF with diverse protein partners (Fig. 1B,C), but what determines which partner is chosen for each occasion? At this point we do not have clear answers to these questions, but several possibilities can be considered. First, CTCF uses its 11 zinc fingers in a combinatorial way (Ohlsson et al., 2001**)** to recognize and bind to a variety of DNA sequences (see below). The discriminate usage of a subset of zinc fingers for DNA binding might create, out of the remaining fingers, specific platforms for interaction with other proteins. A second possible mechanism for the control of partner choice and affinity of the CTCF-partner interaction is the different post-translational modifications of the partner and/or of CTCF itself, which might be used under different cellular circumstances. At least one example has been reported in which post-translational modifications of CTCF affect its interaction with a partner protein – in this case, RNA polymerase II (Pol II) (Chernukhin et al., 2007) (see below for further details).

In this Commentary, we first discuss the key features of CTCF, including its DNA-binding specificity and its role in linking intra- and interchromosomal sites. We next focus our attention on several protein partners of CTCF that are known to have important cellular functions, or that have been very recently identified [CTCF partners that have been identified in proteomic analysis only, such as lamin A/C, importins, topoisomerase II (Topo II) and others (Yusufzai et al., 2004), will not be covered]. In doing so, we will attempt to disentangle the complex knot of CTCF interactions with other proteins, and to understand how these interactions determine the functions of this fascinating protein.

Key characteristics of CTCF

CTCF is a single polypeptide chain of 727 amino acid residues, the secondary structure of which can be subdivided into three distinct domains – an N-terminal region, a central domain containing 11 zinc fingers, and a C-terminal region (reviewed by Ohlsson et al., 2001). The protein sequence is highly conserved among birds and mammals, being 100% identical in the zinc-finger domain. The three domains contain sites for distinct post-translational modifications: the N-terminus is poly(ADP-ribosyl)ated (Yu et al., 2004), whereas the C-terminal domain contains several sites for phosphorylation by casein kinase 2 (Klenova et al., 2001; El-Kady and Klenova, 2005). A recent study reported that CTCF is also modified by SUMOylation (covalent addition of the small ubiquitin-like protein SUMO) at two sites in the polypeptide chain. This modification might contribute to the repressive function of CTCF on the *Myc* P2 promoter (MacPherson et al., 2009). The three distinct domains of CTCF also provide interaction platforms for various proteins (Fig. 1C), including CTCF itself (e.g. Pant et al., 2004; Yusufzai et al., 2004; Ling et al., 2006). The ability of CTCF to dimerize and/or multimerize might underpin its ability to link sites within and between chromosomes (looping and bridging, respectively) (Williams and Flavell, 2008; Zlatanova and Caiafa, 2009) (see also below).

The *CTCF* gene is cell-cycle-regulated, with its expression peaking at S-G2 phase (Klenova et al., 1998). CTCF is characterized by a relatively uniform nuclear distribution in interphase, with prominent binding sites at the periphery of the nucleolus. CTCF also binds to the nuclear matrix, a proteinaceous meshwork in the nucleus that stabilizes nuclear architecture and mechanically supports nuclear processes. This interaction indicates a possible functional connection between CTCF-dependent insulator elements and the nuclear matrix (Dunn et al., 2003). [Insulators are short, specific nucleotide sequences that collaborate with proteins to define boundaries between neighboring, but functionally distinct, genomic domains (Gaszner and Felsenfeld, 2006; Wallace and Felsenfeld,

2007).] The interactions of CTCF with the matrix, as well as with the nucleolus, might occur through the nuclear phosphoprotein nucleophosmin (Yusufzai and Felsenfeld, 2004; Yusufzai et al., 2004) (and see below). CTCF also associates with the centrosomes and the midbody at the end of mitosis, suggesting that it has non-nuclear functions, such as cell-cycle control (Zhang et al., 2004).

DNA-binding specificity and genome-wide distribution of CTCF

CTCF was originally described as a transcriptional repressor of the chicken, mouse and human *Myc* genes (Lobanenkov et al., 1990; Klenova et al., 1993; Filippova et al., 1996). Since then, CTCF-binding sites have been found in numerous genes, and binding of CTCF to these sites has been implicated in complex transcriptional regulation pathways. Early attempts to define a consensus CTCF-binding DNA sequence were unsuccessful and the diversity of identified binding sequences indicated that CTCF had an exceptional degree of flexibility in terms of binding-site recognition. This flexibility was attributed to combinatorial usage of the 11 zinc fingers in the central part of the molecule, and led to the description of CTCF as a 'multivalent' transcription factor (Ohlsson et al., 2001). Recent genome-wide chromatin immunoprecipitation (ChIP) experiments utilized microarrays (ChIP-on-chip) (Kim et al., 2007) or Solexa sequencing technology (ChIP-seq) (Barski et al., 2007) to identify the DNA sequences immunoprecipitated by anti-CTCF antibodies. The studies identified ~14,000 CTCF-binding sites in the human genome, which enabled the derivation of a ~20 bp consensus CTCF-binding sequence (Kim et al., 2007); notably, however, 18% of the sites identified by ChIP experiments did not conform to the consensus sequence, in agreement with earlier observations of CTCF-binding sites on individual genes (Ohlsson et al., 2001). A very similar consensus sequence was simultaneously derived by purely computational approaches in a search for regulatory motifs in conserved non-coding elements in the human genome (Xie et al., 2007). The total number of identified CTCF-binding sites was close to 15,000.

Are there any characteristic features of CTCF distribution that can be gleaned from these genome-wide studies? The CTCF-binding sites correlate with genes but are not close to promoters (Kim et al., 2007; Xie et al., 2007). They often flank groups of genes that are transcriptionally co-regulated, suggesting that the majority of CTCF-binding sites function as insulators. Another recent study identified domains in the human genome that are associated with the nuclear-lamina structure and, more specifically, with lamin B (Guelen et al., 2008). These so-called lamina-associated domains (LADs), which have an average size of ~550 kb, cover 40% of the genome and contain gene-poor regions in a repressive chromatin environment. Computational analysis indicated that 22% of LADs have CTCF-binding sites on one side, and 2% are flanked by two binding sites. The CTCF-binding sites center at 5-10 kb outside the LAD borders; however, these sites of CTCF accumulation do not coincide with the sites of promoter enrichment in these regions, in agreement with Kim et al. (Kim et al., 2007) and Xie et al. (Xie et al., 2007).

Linking intra- and interchromosomal sites

As has been discussed above, CTCF appears to be able to link discrete domains on the same or different chromosomes. An important series of studies has used various modifications of the chromosome conformation capture (3C) technique (Dekker et al., 2002) to identify chromatin regions that contact each other

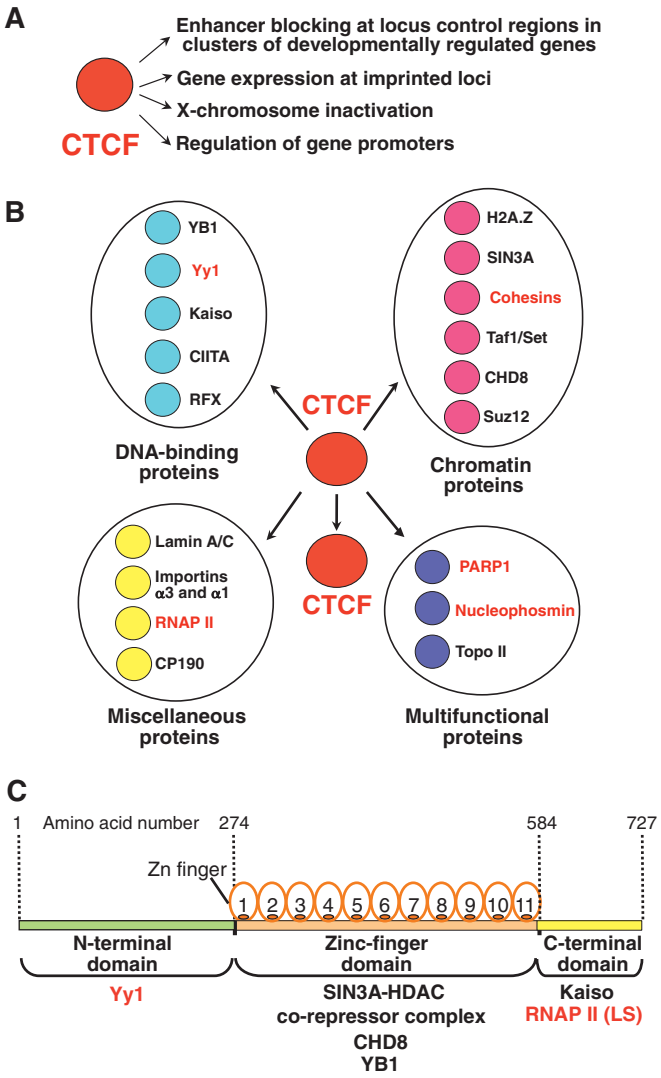

Fig. 1. Cellular functions, protein partners and structure of CTCF. (A) CTCF has important roles in numerous cellular processes. (B) Recognized protein partners of CTCF, broadly grouped according to function. The protein partners that are discussed in detail in this Commentary are highlighted in red. (C) Schematic of CTCF primary structure, showing its three domains, as well as those protein partners whose interactions with CTCF have been mapped to the individual domains. CHD8, chromodomain helicase DNA-binding protein 8; CIITA, MHC class II transactivator; CP190, centrosomal protein 190; H2A.Z, variant Z of histone H2A; LS, large subunit; HDAC, histone deacetylase; PARP1, poly(ADP-ribose) polymerase 1; SIN3A, SIN3 homolog A, transcription regulator (yeast); RFX, regulatory factor X; RNAP II, RNA polymerase II; Suz12, suppressor of zeste 12 homolog (*Drosophila*), Taf1/Set, SET translocation (myeloid leukemia-associated); Topo II, DNA topoisomerase II; YB1, Y-box binding protein 1; Yy1, yin and yang 1.

physically in the nucleus. Ling and co-workers (Ling et al., 2006) studied the transcriptional control of imprinted genes (only one of the two alleles of such genes is expressed in a given cell, and the expression is determined by the parental origin of the allele, i.e. the mother and the father alleles are differentially expressed). They found that the so-called imprinting control region (ICR) that borders and governs the expression of the maternal allele of the *H19* gene (which encodes an RNA molecule of unknown function) specifically interacts with the paternal allele of an intergenic region between two other imprinted genes, *Wsb1* (WD repeat and SOCS-

box-containing protein 1) and *Nf1* (neurofibromin). Notably, the interacting regions are located on two different mouse chromosomes: the *H19* ICR on chromosome 7, and *Wsb1* and *Nf1* on chromosome 11. Importantly, the interaction is dependent on the presence of CTCF and intact CTCF-binding sites. A further study identified 114 unique sequences from all chromosomes that interact with the same *H19* ICR region, with some preference for interchromosomal interactions (Zhao et al., 2006). Imprinted loci are highly represented among the interacting DNA regions, and the pattern of interactions changes during differentiation ex vivo (in embryonic stem cells) and in vivo (when comparing the embryoid body with the neonatal liver). Notably, the physical proximity of sites depends on intact CTCF target sites, implicating CTCF in mediating these interactions [for further examples and discussion, see Zlatanova and Caiafa (Zlatanova and Caiafa, 2009)].

Simonis and colleagues (Simonis et al., 2006) studied the β-globin gene locus in its transcriptionally active (fetal liver) and inactive (fetal brain) state. When active, the locus preferentially interacts with other transcribed loci, whereas the inactive locus prefers to partner with transcriptionally silent regions. This study did not directly address CTCF involvement in bringing genomic loci together; however, such an involvement is to be expected because CTCF is known to have a role in enhancer blocking in the β-globin gene clusters through a mechanism that involves loop formation, i.e. bringing distant DNA regions together (Bell et al., 1999; Farrell et al., 2002; Splinter et al., 2006).

CTCF-interacting proteins – a different partner for each occasion?

The number of proteins recognized to interact with CTCF under specific circumstances is growing steadily and will, undoubtedly, continue to grow. In general, CTCF partners can be divided into several functional groups (Fig. 1B). The group of DNA-binding proteins [transcription factors (activators and/or repressors depending on the context) and cofactors] includes, but is probably not limited to, Y-box-binding protein 1 (YB1) (Chernukhin et al., 2000), Yin and yang 1 (Yy1) (Fig. 2), Kaiso (Defossez et al., 2005), and regulatory factor X (RFX) and MHC class II transactivator (CIITA) (Majumder et al., 2006; Majumder et al., 2008). The second category of partners includes chromatin proteins (both structural proteins and enzymes). Table 1 provides a summary of the most important characteristics of each specific partner, and the main findings concerning the functional significance of the partnership, for the first two groups of interactors. A third group includes important multifunctional proteins, such as poly[ADP-ribose] polymerase 1 (PARP1), nucleophosmin and Topo II. Finally, there are other identified partners that do not belong to any of these groups and will be separately considered as ‘miscellaneous’. In the following subsections, we describe the interactions of CTCF with several partner proteins, and show how these give rise to distinct functions of CTCF. Please note, however, that these cases represent only a few examples of the CTCF-protein interactions that occur at specific genomic loci; in addition, the issue of whether CTCF actually recruits the partner protein in question to the site has not been addressed in most of the examples described.

Yy1 is a CTCF partner with a role in X-chromosome inactivation

Yy1 is a ubiquitous four-zinc-finger transcription factor that has been implicated in biological processes such as embryogenesis, differentiation, cell proliferation and tumorigenesis (Gordon et al., 2006). Homozygous *Yy1* mouse mutants die early in development, whereas heterozygous animals are characterized by severe growth retardation and neurological defects (Gordon et al., 2006). It has been hypothesized that overexpression and/or activation of Yy1 are linked to loss of control of cell proliferation, although the molecular mechanisms remain elusive. Among the numerous potential mechanisms are effects on p53 expression and/or activity (Gordon et al., 2006) and stimulation of PARP1 activity (Griesenbeck et al., 1999). PARP1 stimulation might be of special interest, because PARP1 has been identified as a CTCF interaction partner (Yusufzai et al., 2004) and poly(ADP-ribosyl)ated forms of CTCF have been implicated in the control of transcription of imprinted genes and ribosomal DNA (Yu et al., 2004; Torrano et al., 2006; Caiafa and Zlatanova, 2009) (see below). Vertebrate Yy1 has also been implicated in polycomb group (PcG)-mediated functions because it can repress transcription in *Drosophila* and functionally compensates for loss of its *Drosophila* homologue, PHO (Atchison et al., 2003; Wilkinson et al., 2006). Yy1 recruits the PcG complex to DNA, resulting in methylation of histone H3K27 (Wilkinson et al., 2006); the introduction of methyl groups onto Lys27 in the tail of histone H3 is thought be a mechanism through which PcG proteins repress expression of genes involved in embryonic development.

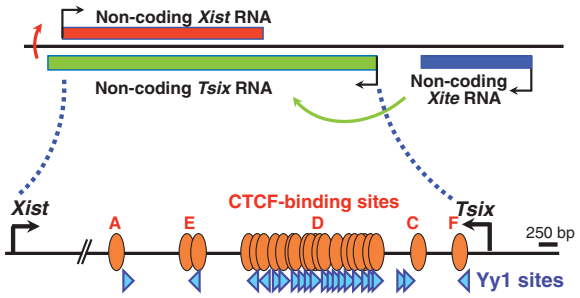

Fig. 2. Yy1 and CTCF collaborate in the regulation of X-chromosome inactivation. The portion of the X-inactivation center that encodes the three non-coding RNA transcripts, *Xite*, *Tsix* and *Xist*, that perform the binary switch function is shown; active *Xist* transcription initiates the actual inactivation of the randomly selected X chromosome (see text). The numerous groups of CTCF-binding sites (A, E, D, C and F) in this region are often paired with Yy1 sites (the orientation of these sites is represented by triangles). The direct interaction of CTCF with Yy1 at these sites is believed to enhance *Xist* transcription. Schematics modified from Ogawa et al. (Ogawa et al., 2008) and Donohoe et al. (Donohoe et al., 2007).

Yy1 has been recently identified as a CTCF cofactor that has a role in X-chromosome inactivation. Although the mechanism still remains unclear, it is worth noting that another CTCF partner, histone variant H2A.Z, has been also implicated in the inactivation process (Donohoe et al., 2007) (Fig. 2). In mammals, gene-dosage compensation between females (XX) and males (XY) occurs through a random inactivation of one of the two female X chromosomes. The inactivation process is complex and occurs through at least three genetically separable stages: (1) ‘counting’ of the X-chromosome-to-autosome ratio to ensure the inactivation of only one of the two X chromosomes; (2) ‘choice’ of the chromosome to be inactivated; and (3) the actual inactivation process, which is initiated by coating the designated inactive chromosome with the non-coding *Xist* RNA (Avner and Heard, 2001; Clerc and Avner, 2006; Erwin and Lee, 2008). CTCF has been implicated in the initial pairing of the two X chromosomes through their X-inactivation centers (Avner and Heard,

Table 1. CTCF protein partners involved in binding to, or modification of, DNA or chromatin

Protein partner	Function	Main observation	Reference
DNA-binding proteins			
YB1	Multifunctional DNA- and RNA-binding factor implicated in regulation of DNA replication, DNA repair, transcription and RNA processing; interacts with Yy1	Co-immunoprecipitates with CTCF in vivo; interacts with CTCF zinc-finger domain; cooperates with CTCF in transcriptional repression of *Myc*	Chernukhin et al., 2000
		CTCF interferes with the binding of YB1 to transcription control elements (variable-number tandem-repeat domains) in intron 2 of the gene encoding the serotonin transporter 5-HTT, which has been implicated in CNS-related disorders	Klenova et al., 2004
Yy1	Zinc-finger transcription factor	Paired CTCF-Yy1 binding sites are highly clustered at the *Tsix* domain of the X-chromosome inactivation center (see text and Fig. 2 for details)	Donohoe et al., 2007
		In transient co-transfection experiments, Yy1 specifically interacts with CTCF (mainly through the CTCF N-terminus) to transactivate *Tsix* (to a greater extent than either protein alone)	
Kaiso	Member of the pox-virus and zinc-finger (POZ) family of zinc-finger transcription factors, which are implicated in development and cancer; possesses dual specificity of DNA binding (binds to methylated CpGs or to the non-methylated sequence TGGCAGGA)	Binds to CTCF bait in yeast two-hybrid screen; interaction is through the CTCF C-domain; binds to the unmethylated consensus sequence close to the CTCF-binding site in the human 5′ β-globin insulator and reduces CTCF enhancer-blocking activity	Defossez et al., 2005; De La Rosa-Velázquez et al., 2007
		Replaces CTCF at the promoter of *RB1*, the gene encoding human retinoblastoma-associated protein (Rb), when the CTCF-binding site becomes methylated; binding of Kaiso results in transcriptional repression of *RB1*	
RFX and CIITA	RFX is a transcription factor that binds to proximal promoters of all MHCII genes (and is required, but not sufficient, for expression); CIITA is a transcriptional co-activator that controls expression by recruiting chromatin remodelers and transcription factors	CTCF directly interacts with both RFX and CIITA, probably forming a trimeric complex; the complex is involved in loop formation between the promoters of the *HLA-DRB1* and *HLA-DQA1* genes and the intergenic element *XL9* (which contains a CTCF-binding site) to allow expression of the genes	Majumder et al., 2006; Majumder et al., 2008
Chromatin proteins			
H2A and H2A.Z	Structural components of nucleosomes; H2A.Z is a non-allelic histone H2A variant that replaces H2A in nucleosomes at specific genome locations (Zlatanova and Thakar, 2008)	Identified as CTCF cofactors by CTCF-affinity chromatography followed by mass-spectrometry analysis	Yusufzai et al., 2004
		Co-immunoprecipitate with CTCF in vivo	Guastafierro et al., 2008
		Co-localize with CTCF genome-wide	Barski et al., 2008
		CTCF positions 20 nucleosomes around H2A-binding sites (genome-wide); these nucleosomes are highly enriched for H2A.Z and 11 post-translational histone modifications	Fu et al., 2008
Suz12	Essential component of polycomb repressor complex 2 (PRC2), which methylates histone H3 at lysine 27	Binds specifically to the maternal allele of promoters P2 and P3 of the repressed *Igf2* allele at the imprinted *Igf2/H19* locus (H3K27 becomes methylated at the maternal allele); Suz12 directly interacts with CTCF both in vivo and in vitro	Li et al., 2008; Han et al., 2008
SIN3A	Transcriptional co-repressor	Binds to CTCF via the zinc-finger domain; recruits histone deacetylase activity	Lutz et al., 2000
CHD8	Member of the chromodomain helicase family, which is implicated in chromatin assembly and control of gene expression	Binds to the CTCF zinc-finger domain used as bait in a yeast two-hybrid screen; associates with known CTCF-binding sites (*H19* ICR, 5′ HS5 of the LCR of β-globin gene cluster, and the promoters of *BRCA1* and *Myc*; knockdown of either CTCF or CHD8 results in loss of ICR insulator activity at luciferase reporter plasmids; CHD8 acts through CTCF at reporter plasmids and the endogenous ICR site; loss of CHD8 induces CpG hypermethylation and histone hypo-acetylation in the vicinity of CTCF-binding sites at *BRCA1* and *Myc* promoters	Ishihara et al., 2006
Taf1/Set	Molecular chaperone; component of the INHAT complex that inhibits histone acetyltransferases	Identified as a CTCF cofactor by CTCF-affinity chromatography followed by mass-spectrometry analysis	Yusufzai et al., 2004
CP190	Centrosome-binding protein that also binds to *Drosophila* polytene chromosomes; essential for viability but not required for cell division	CP190-binding sites significantly overlap with those of CTCF in *Drosophila*; CP190 is required for proper CTCF binding to chromatin; CTCF localizes at the borders of interbands and bands on polytene chromosomes; CP190 directly interacts with CTCF in vivo	Mohan et al., 2007; Gerasimova et al., 2007
Cohesin	Four-subunit complex (Smc1, Smc3, Scc1 and Scc3) that forms a ring-like structure in sister-chromatid cohesion; implicated in proper chromosome segregation and homologous-recombination-dependent DNA-damage repair	Cohesin colocalizes with CTCF at: the control region of the major latency-associated transcript (*LAT*) gene of Kaposi sarcoma-associated herpesvirus (and dissociates upon lytic-cycle induction); ICR of imprinted mouse *Igf2/H19* locus; and the *Myc* promoter	Stedman et al., 2008
		Cohesin colocalizes with CTCF in the human genome; CTCF recruits cohesin to specific sites; cohesin is required for insulator function at *H19* ICR and human β-globin locus at reporter plasmids; cohesin and CTCF are bound to the same (maternal) DNA molecules; controls transcription at the *Igf2/H19* imprinted locus in both G1 and G2 cells (although cohesion does not occur in G1 cells)	Wendt et al., 2008
		Cohesin colocalizes with CTCF in mammalian cells (conventional ChIP and ChiP-on-chip) (70% of all identified cohesin and CTCF sites are co-occupied by both proteins); CTCF recruits cohesin to specific sites; insulator function of cohesin on transfected insulator plasmid is lost by siRNA-mediated depletion of either CTCF or Rad21	Parelho et al., 2008
		Interacts with CTCF at the *Myc* insulator; recruitment of cohesin to chromosomal sites (*Igf2/H19* and *DM* locus) depends on the presence of CTCF; colocalizes (within 1 kb) with CTCF in the human genome (ChIP-on-chip); some chromosomal sites interact exclusively with CTCF or cohesin	Rubio et al., 2008

2001; Clerc and Avner, 2006; Erwin and Lee, 2008), in the 'choice' decision (e.g. Xu et al., 2007), and in the inactivation process itself (Pugacheva et al., 2005). CTCF is also involved in the function of boundary (insulator) elements that separate inactivated genes from rare 'escapee' genes that remain transcriptionally active in the context of the inactive X chromosome (Filippova et al., 2005). The interactions of Yy1 and CTCF are described in more detail in Table 1.

Next, we describe the role of CTCF in X-chromosome inactivation in more detail. The physical map of the region that specifies the sequences of the three non-coding RNAs involved in the inactivation process is presented in Fig. 2. On the future active X chromosome, *Xite* (X-inactivation intergenic transcription element) prolongs the antisense transcription of *Tsix* [X (inactive)-specific transcript, antisense], which in turn blocks transcription of *Xist* [X (inactive)-specific transcript] (Fig. 2); both CTCF and Yy1 transactivate *Tsix*. On the future inactive X chromosome, repression of *Xite* downregulates *Tsix* transcription, which in turn induces *Xist* transcription to initiate the inactivation process. In mouse cells, the *Xist-Tsix* region is characterized by the presence of ~40 potential CTCF-binding sites, which are frequently paired with binding sites for Yy1 (Donohoe et al., 2007) (Fig. 2). CTCF directly interacts with Yy1, as shown in co-immunoprecipitation experiments; the high-affinity interaction between the two proteins involves mainly the N-terminus of CTCF (Donohoe et al., 2007). Finally, transient cotransfection experiments indicate that CTCF and Yy1 together confer higher transactivation on *Tsix* than either protein alone (Donohoe et al., 2007). The physical and functional interaction of CTCF with Yy1 during X-chromosome inactivation provides a clear example of how a specific function of CTCF is mediated by a specific protein partner.

Cohesin partners CTCF in gene regulation

The cohesin complex has a central role in holding the two sister chromatids in close contact from the time of DNA replication in S phase to the time of their separation at the onset of mitotic anaphase (reviewed by Hirano and Hirano, 2006; Hirano, 2006). Cohesin function is essential for genome stability and repair; several human developmental disorders, such as Cornelia de Lange syndrome and Robert's syndrome, are associated with mutations in cohesin components or the machinery that loads cohesin on chromatids.

The cohesin complex comprises four subunits; Smc1 and Smc3 are members of the structural maintenance of chromosomes (SMC) protein family, whereas Scc1 and Scc3 (subunit of the cohesin complex 1 and 3) are thought to participate in the formation of a ring structure around the two chromatids (Fig. 3; and see below). Two other non-SMC proteins, Scc2 and Scc4, are required in mammals to load the cohesin complex onto DNA. SMC proteins are large polypeptides of very unusual three-dimensional organization, in which two long α-helices fold back on themselves in an antiparallel orientation to form a rigid coiled-coil domain that has a hinge domain at one end and an ATP-binding 'head' domain at the other (Fig. 3A). Two SMC monomers dimerize at their hinge region to produce long V-shaped molecules. These dimers can form several alternative structures – rings, filaments and rosettes – through intra- and intermolecular interactions. The cohesins are proposed to form ring structures around the two sister chromatids (Haering et al., 2002).

Recently, a cohesion-independent function of cohesins has been recognized in yeast, *Drosophila* and mammals: they have been detected in post-mitotic cells that lack chromatid cohesion and have been implicated in gene regulation (for reviews, see Göndör and Ohlsson, 2008; Peric-Hupkes and van Steensel, 2008; Uhlmann, 2008; Gause et al., 2008). Four recent papers have reported a strong functional connection between cohesins and CTCF (Table 1). First, cohesin proteins and CTCF colocalize both at specific loci (Stedman et al., 2008; Rubio et al., 2008), including the *Myc* insulator element (MINE) (Gombert et al., 2003) (Fig. 3B) and genome-wide (Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008). Second, CTCF recruits cohesin to specific sites, including the *DM1* locus, which has a CTG repeat that is expanded in individuals with myotonic dystrophy (Fig. 3C) (Rubio et al., 2008; Cho et al., 2005). Third, in transient transfection experiments, the activity of insulator elements depends on the presence of cohesin proteins (Parelho et al., 2008; Wendt et al., 2008) (such effects have yet to be demonstrated on endogenous sites).

These studies, exciting as they are, raise a plethora of important questions. For example, what are the molecular interactions that are responsible for the colocalization of CTCF and cohesin? Despite the fact that ~70% of all sites identified as CTCF- and cohesin-binding sites bind to both proteins (Parelho et al., 2008), it is clear that there are sites occupied exclusively by CTCF or cohesin (Rubio et al., 2008). Moreover, downregulation of CTCF does not interfere with mitosis (Parelho et al., 2008; Wendt et al., 2008), suggesting that the cohesion function of cohesin is independent of CTCF. A second question is whether the structure of cohesin is different at CTCF-dependent and CTCF-independent binding sites. Fluorescence recovery after photobleaching (FRAP) experiments suggest that this might be the case; they indicate the existence of two pools of cohesin at interphase (an immobile fraction that is irreversibly bound to chromatin and a dynamic fraction) (Gerlich et al., 2006). The existence of the two distinct cohesin pools is consistent with available biochemical data (Hirano and Hirano, 2006), which suggest the existence of two forms of chromatin-bound cohesin: the ring form that embraces two DNA helices tightly and steadily without interacting directly with DNA, and a less tightly bound form that interacts with DNA in a more conventional manner. The second structure might require other DNA-binding proteins, such as CTCF. We propose that the ring structure is involved in cohesion, whereas the conventional structure participates in gene regulation. Whether long-range chromosomal interactions (loops) are involved in gene regulation through CTCF and cohesin also remains to be directly addressed.

Thus, the interactions between CTCF and cohesin provide another important example of how different CTCF partners may underlie distinct CTCF functions. The cohesin complex should clearly be considered as an interaction partner that mediates the involvement of CTCF in gene regulation.

PARP1 partners CTCF in DNA methylation

Poly(ADP-ribose) polymerases (PARPs) are enzymes that catalyze the formation of poly(ADP-ribose) chains (PARs) on chromatin proteins, including themselves (D'Amours et al., 1999; Schreiber et al., 2006; Kraus, 2008). PARPs use the coenzyme NAD^+ as a source of ADP-ribose moieties to synthesize protein-bound polymers of variable size (ranging from 2 to more than 200 units) and structural complexity (linear or branched); these polymers introduce negative charges onto the acceptor proteins, thus affecting their interactions with DNA and/or other proteins. The intracellular levels of PARs are under tight control; this involves dynamic formation of polymers by members of the PARP family

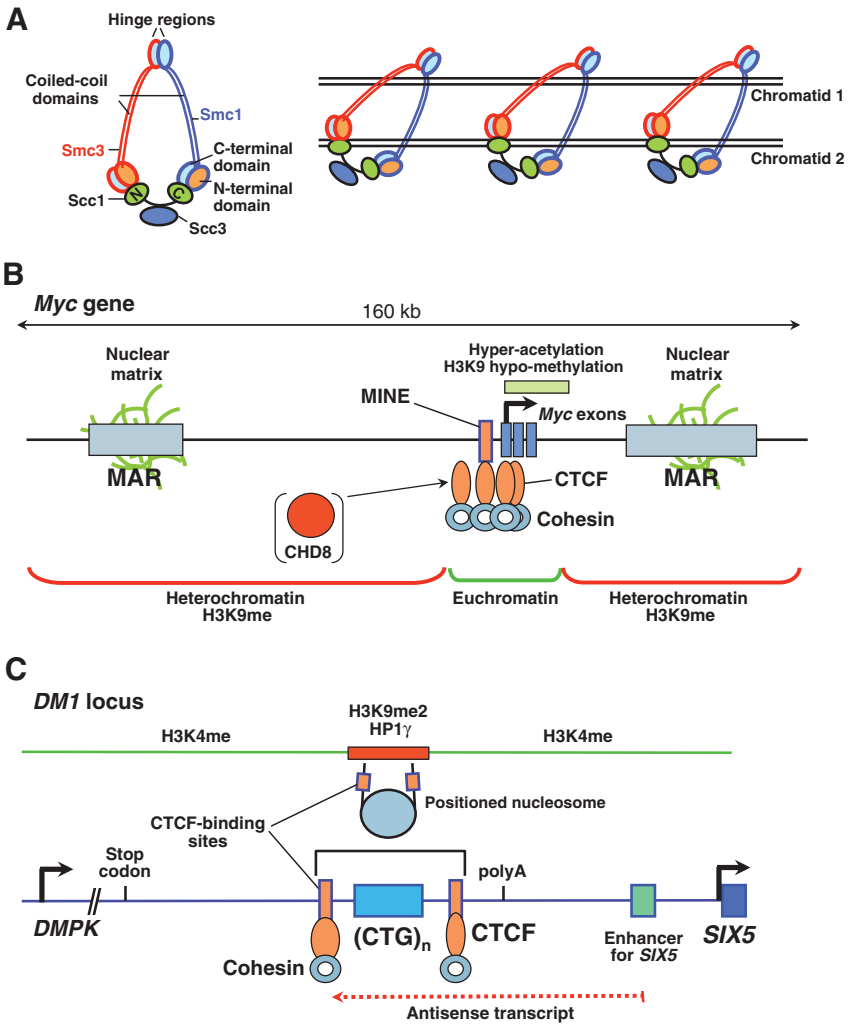

Fig. 3. A role for cohesin and CTCF in gene regulation. (A) The schematic on the left indicates the overall structure and composition of cohesin, showing the long coiled-coil domains in each monomer of Smc1 and Smc3, the hinge regions that connect the two monomers in the heterodimeric structure, and the two other proteins in the complex (Scc1 and Scc3) that close the cohesin ring. The schematic on the right shows the ring model of cohesin structure, in which cohesin embraces sister chromatids in cohesion [redrawn from Hirano (Hirano, 2006)]. (B) CTCF and cohesin colocalize at several CTCF-binding sites, including the *Myc* insulator element (MINE) (Gombert et al., 2003). CTCF is constitutively bound at MINE and at the *Myc* promoter, and binding is independent of the transcriptional status of the gene. The *Myc* gene and its insulator are embedded in a large (~160 kb) domain that is flanked by matrix-attachment regions (MARs) and is devoid of other expressed genes; together, they constitute a euchromatic region embedded within a heterochromatic environment [this might be representative of a more general pattern that was recently recognized on mammalian chromosome arms (Regha et al., 2007) of active chromatin interspersed with repressive chromatin]. The CTCF-binding sites at MINE and the *Myc* promoter also bind to cohesin (Rubio et al., 2008; Stedman et al., 2008). Binding of the chromatin remodeler CHD8 to this region (see bracket) suggests that the chromatin structure in the region is actively altered (Ishihara et al., 2006). (C) The *DM1* locus (which contains *DMPK*, the gene encoding myotonic dystrophy protein kinase), showing the position of the CTG repeat in the 3′ UTR of *DMPK* that is expanded in individuals with myotonic dystrophy. The repeat is flanked by two CTCF-binding sites that are occupied by CTCF. In healthy individuals, the repeat is organized in a single positioned nucleosome (a nucleosome in which the histone octamer occupies a specific sequence). This strict positioning of the single nucleosome over the CTG repeat places the CTCF sites in the DNA-linker regions upstream and downstream of the nucleosome. The chromatin structure of the positioned nucleosome is highly heterochromatic [histone H3 is dimethylated at lysine 9 (H3K9me2)], but the rest of the region is characterized by the presence of 'active' histone modifications [histone H3 is methylated at lysine 4 (H3K4me)]. CTCF restricts the length of the antisense transcript, which limits heterochromatin formation to only the positioned nucleosome. In individuals with myotonic dystrophy, expansion of the CTG repeats is associated with loss of CTCF binding and conversion of the entire region to heterochromatin. According to Rubio et al. (Rubio et al., 2008), the CTCF-binding sites on the human *DM1* locus (integrated in mouse cells) are simultaneously bound by CTCF and cohesin, and binding of cohesin directly depends on the presence of CTCF. Schematic based on Filippova et al. (Filippova et al., 2001) and Cho et al. (Cho et al., 2005). HP1γ, heterochromatin protein 1γ.

(Ame et al., 2004) and their removal by poly(ADP-ribose) glycohydrolase (PARG) (Bonicalzi et al., 2005; Caiafa et al., 2008).

Heteromodification and automodification are the two processes through which PARPs introduce covalently bound ADP-ribose polymers onto other proteins or onto themselves, respectively. Automodification of PARPs is generally activated by nicks on DNA. PAR polymers on PARP1, which are attached at up to 28 sites in the automodification domain, are usually very long (up to 200 ADP-ribose units) and heavily branched (Juarez-Salinas et al., 1982). In addition, PARs (both protein-free and covalently linked to proteins) are capable of strong non-covalent binding (Malanga and Althaus, 2005) to specific proteins, the activity of which is then modulated by the bound polymers.

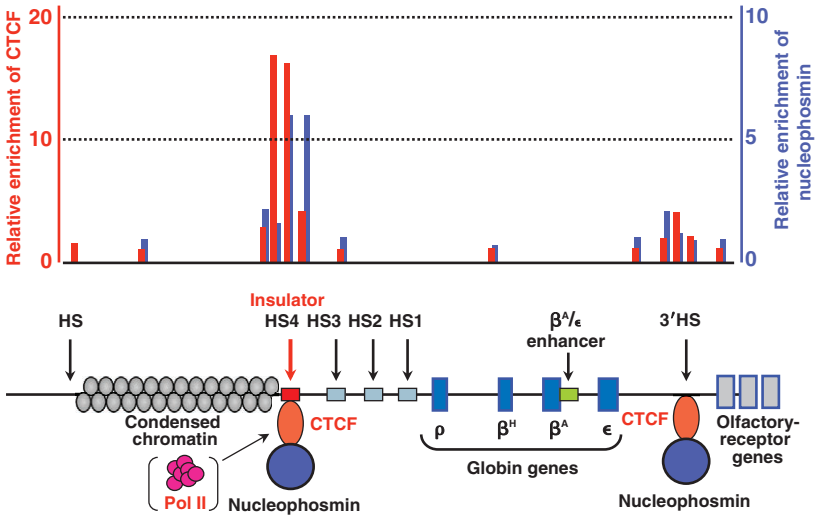

Fig. 4. Nucleophosmin and CTCF associate with both 5′ and 3′ insulator elements in the chicken β-globin gene cluster in vivo. The schematic at the bottom depicts the developmentally regulated β-globin gene cluster and its locus control region (LCR), which encompasses DNase-I-hypersensitive sites 1-3 (HS1-HS3) and the β^A/ε enhancer (Gaszner and Felsenfeld, 2006). The domain is flanked by a region of highly compacted chromatin at the 5′ end and a cluster of genes encoding olfactory receptors at the 3′ end. Two further DNase-I-hypersensitive sites, HS4 and 3′HS, possess enhancer-blocking and insulator activities. 3′HS prevents the β^A/ε enhancer from activating the olfactory-receptor genes, and HS4 acts as both an insulator, to prevent spreading of heterochromatin into the gene cluster, and an enhancer-blocker, to prevent the enhancer located 5′ of the condensed chromatin region from activating the globin genes (enhancer-blocking insulators are effective only when situated between a promoter and an enhancer). Nucleophosmin binds to both HS4 and 3′HS, as shown in the schematic and demonstrated by the ChIP data presented at the top of the figure [modified from Yusufzai et al. (Yusufzai et al., 2004)]. The RNA Pol II complex shown in brackets (Pol II) has been shown by ChIP analysis to localize to the HS4 site (Chernukhin et al., 2007) (see discussion on Pol II in text).

A PARP has been identified among the partners of CTCF in a proteomic search carried out on purified CTCF complexes (Yusufzai et al., 2004). Yu and colleagues (Yu et al., 2004) demonstrated that CTCF undergoes covalent poly(ADP-ribosyl)ation in the N-terminal domain. These authors found that the control of gene imprinting by CTCF is lost upon inhibition of PARP activity, and therefore suggested that PARylated CTCF is directly involved in the control of imprinting. PARylated CTCF has also been implicated in the control of ribosomal gene expression (Torrano et al., 2006; Caiafa and Zlatanova, 2009). Importantly, it has been recently shown that transient ectopic overexpression of CTCF induces PAR accumulation, PARP1 expression and PARylation of CTCF (Guastafierro et al., 2008). In vitro data from this paper have shown that CTCF can activate automodification of PARP1, even in the absence of nicked DNA; this finding is of great interest, because so far a burst of PARylation of PARP1 has generally been found only following introduction of DNA strand breaks. The persistence of high PAR levels over time affects the DNA methylation machinery: DNA-methyltransferase activity is inhibited, with the consequence that the genome becomes diffusely hypomethylated (Caiafa et al., 2008). Thus, the data of Guastafierro and co-workers (Guastafierro et al., 2008) provide, for the first time, evidence that CTCF is involved in the crosstalk between PARylation and DNA methylation, through its activation of PARP1 (which, in turn, leads to inhibition of DNA methylation) (Reale et al., 2005).

Nucleophosmin is a CTCF partner at insulator sites

Nucleophosmin is an abundant nuclear-matrix phosphoprotein, a large fraction of which is localized to the peripheral region of the nucleolus. It has been implicated in embryonic development and maintenance of genomic stability, mainly through its role in centrosome duplication (Grisendi et al., 2005). At the molecular level, nucleophosmin mediates diverse functions, including rDNA transcription, pre-ribosomal RNA processing, mRNA polyadenylation, and the stress response. It also participates in transport functions, chaperoning ribosomal subunits and/or histones from the cytoplasm to the nucleus and nucleoli. A recent study of the role of nucleophosmin in transcriptional regulation of rDNA has indicated that nucleophosmin is associated with the gene locus, maintaining an open chromatin conformation over the active copies of the rRNA genes by removing histones from the promoter (Murano et al., 2008).

Nucleophosmin was identified as a CTCF partner in a proteomic search (Yusufzai et al., 2004), and was the only protein in the soluble CTCF complex that was present in stoichiometric amounts. ChIP analysis of the two known insulator sites that flank the chicken β-globin gene locus confirmed the presence of CTCF at these sites. Remarkably, nucleophosmin was also present at both sites (Fig. 4) (Yusufzai et al., 2004). In human cell lines carrying multiple integrated copies of the chicken HS4 insulator (one of the insulators upstream of the β-globin gene locus), the insulator sites were preferentially localized to the nuclear periphery. As in the case of the endogenous insulator sites at the β-globin gene locus (see above), CTCF colocalized with nucleophosmin at these integrated insulator sites; importantly, the peripheral nucleolar localization of insulator sites was dependent on the integrity of CTCF-binding sites. Thus, it was suggested that insulators are recruited to the periphery of the nucleolus through the strong interaction of CTCF with nucleophosmin (Yusufzai et al., 2004). It should be noted that these data concern only the relatively small portion of CTCF that is located in the nucleolus; a large fraction of CTCF is not bound to the nucleolus, and might not be associated with nucleophosmin (Yusufzai et al., 2004).

Finally, a recent study focused on chromosome translocations involving the immunoglobulin heavy chain (IgH) gene locus in certain cancer cells (Liu et al., 2008). Interestingly, CTCF and nucleophosmin colocalized at the 3′ regulatory elements of the IgH

gene locus only in cells carrying the chromosome translocation; moreover, the cells could be growth arrested by nucleophosmin short hairpin RNA. The exact molecular mechanism behind these observations awaits further research.

The studies described here provide evidence that the insulator function of CTCF is mediated through its specific tethering to subnuclear sites through its interactions with nucleophosmin. Thus, the insulator function of CTCF – similar to its functions in X-chromosome inactivation, gene regulation and DNA methylation – might require its interaction with a partner protein specific to that function.

Is RNA polymerase II a CTCF partner in transcriptional regulation?

The function of CTCF in transcriptional regulation is not well understood. However, a recent report has identified direct interactions between CTCF and the large subunit of Pol II (Chernukhin et al., 2007); we will discuss this paper in detail, as it contains data of potential relevance to the role of CTCF in transcriptional regulation.

In vitro, CTCF interacts equally well with the hypophosphorylated and the hyperphosphorylated forms of Pol II, which are known to be involved in transcription initiation and elongation, respectively (Chernukhin et al., 2007). In vivo, however, CTCF exhibits a significant preference for interaction with the hypophosphorylated Pol II form. This interaction is mediated by the C-terminal domain of CTCF (Fig. 1C), which contains the sites for phosphorylation of CTCF (Klenova et al., 2001; El-Kady and Klenova, 2005). Preliminary data (Chernukhin et al., 2007) indicate that in-vitro-phosphorylated CTCF has a lower affinity for Pol II, suggesting that the CTCF–Pol-II interaction might be subject to regulation by CTCF phosphorylation.

In an attempt to gain insight into the functional significance of the reported CTCF–Pol-II interaction, serial ChIP analysis (using anti-CTCF antibodies, followed by anti-Pol II antibodies, as bait) was used to interrogate the in vivo presence of the CTCF–Pol-II complex on the β-globin insulator (see above) (Chernukhin et al., 2007). Interestingly, CTCF colocalizes with Pol II at the insulator only in proliferating chicken erythroblasts that do not express the globin genes. In differentiated cells that transcribe two of the four globin genes in the cluster, the association of both proteins with the insulator is lost. The mechanisms behind these events remain to be determined. Further experiments in human choriocarcinoma cells transfected with wild-type or mutated *H19* ICR (see above) demonstrated that the binding of Pol II to the ICR requires functional CTCF target sites. Finally, a single CTCF-binding site fused to a promoterless luciferase reporter gene conferred transcriptional activity on the gene in stably integrated constructs. This observation suggested that CTCF is a functional equivalent of TATA-box-binding protein (TBP), and thereby allows accurate transcription initiation at some promoters. This is certainly an interesting notion that deserves to be directly addressed in further experiments.

ChIP-on-chip experiments using a previously constructed library of CTCF-binding sites from mouse fetal liver (Mukhopadhyay et al., 2004) were used to identify sites that are co-occupied by CTCF and Pol II in proliferating and resting NIH 3T3 cells (Chernukhin et al., 2007). Only about 10% of the CTCF sites represented on the microarray interacted with Pol II. Of note, 15 out of the 26 sequences that bound to both CTCF and Pol II were not present in the mouse genome database, which contains almost exclusively euchromatic sequences. Thus, CTCF–Pol-II binding probably also occurs at heterochromatic sequences. Finally, the protein complex was also identified in intergenic regions that are 1.5-15 kb from the nearest gene. Chernukhin and colleagues (Chernukhin et al., 2007) suggest that the CTCF–Pol-II complexes at these sites remain intact until a signal for the release of Pol II is received; the released Pol II then initiates transcription of the neighboring genes from cryptic promoters.

An earlier study that is relevant to Pol-II- and CTCF-mediated insulator function showed that the presence of the chicken insulator HS4 on chromatinized episomes (ectopic, unintegrated DNA constructs that acquire characteristics of chromatin organization in the host cell) in human cells leads to accumulation of Pol II at the enhancer in the β-globin gene locus control region (Zhao and Dean, 2004). This suggested that, as part of its insulator function, CTCF blocks the transfer of Pol II from the enhancer to the promoter. Whether and how these observations relate to the more recent data (Chernukhin et al., 2007) remains to be seen.

More recently, a possible link between CTCF binding and Pol II occupancy was revealed in a genome-wide study (Barski et al., 2007), in which a tantalizing high-resolution profiling of histone methylation patterns in the human genome was undertaken. In addition to mapping 20 histone lysine and arginine methylations, the authors addressed the genome-wide localization patterns of Pol II, histone H2A.Z (see Table 1) and CTCF. Out of the ~20,000 CTCF-binding sites, more than 6000 were in transcribed regions. Unfortunately, the CTCF sites that lie close to Pol II sites were excluded from further analysis to avoid complications in the interpretation of the methylation data, which was the main objective of that study.

The picture that emerges from the study by Chernukhin and colleagues (Chernukhin et al., 2007) is extremely complex; the authors suggest several possible functions of the CTCF–Pol-II complex that are context dependent. It is clear that numerous new questions (concerning the mechanism of a possible TBP-like function for CTCF, the presence and distribution of CTCF–Pol-II complexes at different genomic regions, etc.) arise from this study, and that significant experimental effort will be required to address them.

The CTCF-partner network

Above, we have presented and discussed evidence that connects CTCF with individual protein partners, particularly Yy1, cohesin, PARP1, nucleophosmin and Pol II. We have pointed out that the interactions of CTCF with each protein partner occur in a specific biological context. However, it has not escaped our attention that some of the partners are known to interact with each other, thus creating a rather complex network (Fig. 5). For example, nucleophosmin is a recognized partner of PARP1 (Meder et al., 2005), and PARP1 interacts with Yy1 (Oei and Shi, 2001a; Oei and Shi, 2001b). In addition, Yy1 directly interacts with another recognized CTCF partner, YB1 (Chernukhin et al., 2000; Li et al., 1997). It is clear that more research is needed to identify the possible protein interactions in the CTCF network, and to understand the biological contexts in which they work.

Concluding remarks

The data discussed in this Commentary show that CTCF possesses extreme flexibility, not only in terms of the diversity of its binding sites but also with respect to its numerous binding partners. It seems that CTCF performs its numerous functions by using different

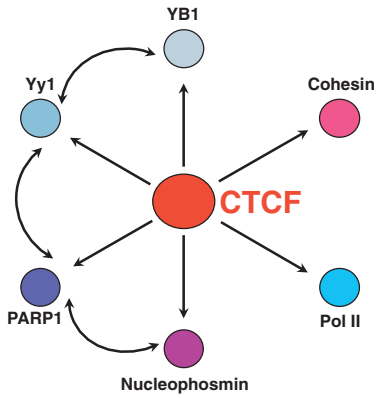

Fig. 5. Summary of the CTCF protein interaction network from the interactions discussed in this Commentary. The arrows connecting individual partners show recognized interactions. Thus, for example, both Yy1 and CTCF upregulate PARP1 activity; PARP1 and nucleophosmin interact directly, which might contribute to the inhibitory effect of CTCF on ribosomal gene transcription.

binding partners in different biological contexts. Two points deserve special mention. First, even with one and the same partner, CTCF is obviously performing a multiplicity of (sometimes seemingly antagonistic) functions. The CTCF–Pol-II interaction might provide a good example of such functional diversity, because such complexes might perform different functions depending on whether they are located in euchromatin or heterochromatic regions. Second, the various partners seem to interact with each other directly or indirectly, which is likely to contribute to the fine-tuning of CTCF function (Fig. 5). There is no doubt that new CTCF protein partners will be identified in the future; they will probably endow CTCF with distinct functions in distinct biological contexts, as the ones that are already recognized appear to do. Will we ever be able to understand this complexity? Is 'divide and rule' the key to success in nature, as well as in society?

J.Z. is supported in part by NSF grant 0504239; P.C. is partially financed by Ministero della Salute, Italy.

References

Ame, J. C., Spenlehauer, C. and de Murcia, G. (2004). The PARP superfamily. *BioEssays* **26**, 882-893.
Atchison, L., Ghias, A., Wilkinson, F., Bonini, N. and Atchison, M. L. (2003). Transcription factor YY1 functions as a PcG protein *in vivo*. *EMBO J.* **22**, 1347-1358.
Avner, P. and Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. *Nat. Rev. Genet.* **2**, 59-67.
Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837.
Bell, A. C., West, A. G. and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387-396.
Bonicalzi, M. E., Haince, J. F., Droit, A. and Poirier, G. G. (2005). Regulation of poly(ADP-ribose) metabolism by poly(ADP-ribose) glycohydrolase: where and when? *Cell Mol. Life Sci.* **62**, 739-750.
Caiafa, P. and Zlatanova, J. (2009). CCCTC-binding factor meets poly(ADP-ribose) polymerase-1. *J. Cell Physiol.* **219**, 265-270.
Caiafa, P., Guastafierro, T. and Zampieri, M. (2008). Epigenetics: poly(ADP-ribosyl)ation of PARP-1 regulates genomic methylation patterns. *FASEB J.* **23**, 672-678.
Chernukhin, I. V., Shamsuddin, S., Robinson, A. F., Carne, A. F., Paul, A., El-Kady, A. I., Lobanenkov, V. V. and Klenova, E. M. (2000). Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *J. Biol. Chem.* **275**, 29915-29921.
Chernukhin, I., Shamsuddin, S., Kang, S. Y., Bergstrom, R., Kwon, Y. W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D. et al. (2007). CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol. Cell. Biol.* **27**, 1631-1648.
Cho, D. H., Thienes, C. P., Mahoney, S. E., Analau, E., Filippova, G. N. and Tapscott, S. J. (2005). Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20**, 483-489.
Clerc, P. and Avner, P. (2006). Random X-chromosome inactivation: skewing lessons for mice and men. *Curr. Opin. Genet. Dev.* **16**, 246-253.
D'Amours, D., Desnoyers, S., D'Silva, I. and Poirier, G. G. (1999). Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochem. J.* **342**, 249-268.
De La Rosa-Velazquez, I. A., Rincon-Arano, H., Benitez-Bribiesca, L. and Recillas-Targa, F. (2007). Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res.* **67**, 2577-2585.
Defossez, P. A., Kelly, K. F., Filion, G. J., Perez-Torrado, R., Magdinier, F., Menoni, H., Nordgaard, C. L., Daniel, J. M. and Gilson, E. (2005). The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J. Biol. Chem.* **280**, 43017-43023.
Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306-1311.
Donohoe, M. E., Zhang, L. F., Xu, N., Shi, Y. and Lee, J. T. (2007). Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol. Cell* **25**, 43-56.
Dunn, K. L., Zhao, H. and Davie, J. R. (2003). The insulator binding protein CTCF associates with the nuclear matrix. *Exp. Cell Res.* **288**, 218-223.
El-Kady, A. and Klenova, E. (2005). Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. *FEBS Lett.* **579**, 1424-1434.
Erwin, J. A. and Lee, J. T. (2008). New twists in X-chromosome inactivation. *Curr. Opin. Cell Biol.* **20**, 349-355.
Farrell, C. M., West, A. G. and Felsenfeld, G. (2002). Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol. Cell. Biol.* **22**, 3820-3831.
Filippova, G. N. (2008). Genetics and epigenetics of the multifunctional protein CTCF. *Curr. Top. Dev. Biol.* **80**, 337-360.
Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P. E., Collins, S. J. and Lobanenkov, V. V. (1996). 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.* **16**, 2802-2813.
Filippova, G. N., Thienes, C. P., Penn, B. H., Cho, D. H., Hu, Y. J., Moore, J. M., Klesert, T. R., Lobanenkov, V. V. and Tapscott, S. J. (2001). CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat. Genet.* **28**, 335-343.
Filippova, G. N., Cheng, M. K., Moore, J. M., Truong, J. P., Hu, Y. J., Nguyen, D. K., Tsuchiya, K. D. and Disteche, C. M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev. Cell* **8**, 31-42.
Fu, Y., Sinha, M., Peterson, C. L. and Weng, Z. (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4**, e1000138.
Gaszner, M. and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7**, 703-713.
Gause, M., Schaaf, C. A. and Dorsett, D. (2008). Cohesin and CTCF: cooperating to control chromosome conformation? *BioEssays* **30**, 715-718.
Gerasimova, T. I., Lei, E. P., Bushey, A. M. and Corces, V. G. (2007). Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. *Mol. Cell* **28**, 761-772.
Gerlich, D., Koch, B., Dupeux, F., Peters, J. M. and Ellenberg, J. (2006). Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. *Curr. Biol.* **16**, 1571-1578.
Gombert, W. M., Farris, S. D., Rubio, E. D., Morey-Rosler, K. M., Schubach, W. H. and Krumm, A. (2003). The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain. *Mol. Cell. Biol.* **23**, 9338-9348.
Göndör, A. and Ohlsson, R. (2008). Chromatin insulators and cohesins. *EMBO Rep.* **9**, 327-329.
Gordon, S., Akopyan, G., Garban, H. and Bonavida, B. (2006). Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* **25**, 1125-1142.
Griesenbeck, J., Ziegler, M., Tomilin, N., Schweiger, M. and Oei, S. L. (1999). Stimulation of the catalytic activity of poly(ADP-ribosyl) transferase by transcription factor Yin Yang 1. *FEBS Lett.* **443**, 20-24.
Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K. and Pandolfi, P. P. (2005). Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* **437**, 147-153.
Guastafierro, T., Cecchinelli, B., Zampieri, M., Reale, A., Riggio, G., Sthandier, O., Zupi, G., Calabrese, L. and Caiafa, P. (2008). CTCF activates PARP-1 affecting DNA methylation machinery. *J. Biol. Chem.* **283**, 21873-21880.
Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., Eussen, B. H., de Klein, A., Wessels, L., de Laat, W. et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-951.
Haering, C. H., Lowe, J., Hochwagen, A. and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**, 773-788.
Han, L., Lee, D. H. and Szabo, P. E. (2008). CTCF is the master organizer of domain-wide allele-specific chromatin at the H19/Igf2 imprinted region. *Mol. Cell. Biol.* **28**, 1124-1135.
Hirano, M. and Hirano, T. (2006). Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions. *Mol. Cell* **21**, 175-186.
Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. *Nat. Rev. Mol. Cell. Biol.* **7**, 311-322.

Ishihara, K., Oshimura, M. and Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol. Cell* **23**, 733-742.
Juarez-Salinas, H., Levi, V., Jacobson, E. L. and Jacobson, M. K. (1982). Poly(ADP-ribose) has a branched structure *in vivo*. *J. Biol. Chem.* **257**, 607-609.
Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V. and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231-1245.
Klenova, E. M., Nicolas, R. H., Paterson, H. F., Carne, A. F., Heath, C. M., Goodwin, G. H., Neiman, P. E. and Lobanenkov, V. V. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell. Biol.* **13**, 7612-7624.
Klenova, E. M., Fagerlie, S., Filippova, G. N., Kretzner, L., Goodwin, G. H., Loring, G., Neiman, P. E. and Lobanenkov, V. V. (1998). Characterization of the chicken CTCF genomic locus, and initial study of the cell cycle-regulated promoter of the gene. *J. Biol. Chem.* **273**, 26571-26579.
Klenova, E. M., Chernukhin, I. V., El-Kady, A., Lee, R. E., Pugacheva, E. M., Loukinov, D. I., Goodwin, G. H., Delgado, D., Filippova, G. N., Leon, J. et al. (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol. Cell. Biol.* **21**, 2221-2234.
Klenova, E., Scott, A. C., Roberts, J., Shamsuddin, S., Lovejoy, E. A., Bergmann, S., Bubb, V. J., Royer, H. D. and Quinn, J. P. (2004). YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. *J. Neurosci.* **24**, 5966-5973.
Kraus, W. L. (2008). Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation. *Curr. Opin. Cell Biol.* **20**, 294-302.
Li, T., Hu, J. F., Qiu, X., Ling, J., Chen, H., Wang, S., Hou, A., Vu, T. H. and Hoffman, A. R. (2008). CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol. Cell. Biol.* **28**, 6473-6482.
Li, W. W., Hsiung, Y., Wong, V., Galvin, K., Zhou, Y., Shi, Y. and Lee, A. S. (1997). Suppression of grp78 core promoter element-mediated stress induction by the dbpA and dbpB (YB-1) cold shock domain proteins. *Mol. Cell. Biol.* **17**, 61-68.
Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., Cherry, A. M. and Hoffman, A. R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312**, 269-272.
Liu, H., Huang, J., Wang, J., Jiang, S., Bailey, A. S., Goldman, D. C., Welcker, M., Bedell, V., Slovak, M. L., Clurman, B. et al. (2008). Transvection mediated by the translocated cyclin D1 locus in mantle cell lymphoma. *J. Exp. Med.* **205**, 1843-1858.
Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V. and Goodwin, G. H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-myc gene. *Oncogene* **5**, 1743-1753.
Lutz, M., Burke, L. J., Barreto, G., Goeman, F., Greb, H., Arnold, R., Schultheiss, H., Brehm, A., Kouzarides, T., Lobanenkov, V. et al. (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* **28**, 1707-1713.
MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M. and Sadowski, P. D. (2009). The CTCF insulator protein is posttranslationally modified by SUMO. *Mol. Cell. Biol.* **29**, 714-725.
Majumder, P., Gomez, J. A. and Boss, J. M. (2006). The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element. *J. Biol. Chem.* **281**, 18435-18443.
Majumder, P., Gomez, J. A., Chadwick, B. P. and Boss, J. M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* **205**, 785-798.
Malanga, M. and Althaus, F. R. (2005). The role of poly(ADP-ribose) in the DNA damage signaling network. *Biochem. Cell Biol.* **83**, 354-364.
Meder, V. S., Boeglin, M., de Murcia, G. and Schreiber, V. (2005). PARP-1 and PARP-2 interact with nucleophosmin/B23 and accumulate in transcriptionally active nucleoli. *J. Cell Sci.* **118**, 211-222.
Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinl, N., Bardenhagen, I., Leers, J., White, R. A., Renkawitz-Pohl, R., Saumweber, H. et al. (2007). The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *EMBO J.* **26**, 4203-4214.
Mukhopadhyay, R., Yu, W., Whitehead, J., Xu, J., Lezcano, M., Pack, S., Kanduri, C., Kanduri, M., Ginjala, V., Vostrov, A. et al. (2004). The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res.* **14**, 1594-1602.
Murano, K., Okuwaki, M., Hisaoka, M. and Nagata, K. (2008). Transcription regulation of the rRNA gene by a multifunctional nucleolar protein, B23/nucleophosmin, through its histone chaperone activity. *Mol. Cell. Biol.* **28**, 3114-3126.
Oei, S. L. and Shi, Y. (2001a). Transcription factor Yin Yang 1 stimulates poly(ADP-ribosyl)ation and DNA repair. *Biochem. Biophys. Res. Commun.* **284**, 450-454.
Oei, S. L. and Shi, Y. (2001b). Poly(ADP-ribosyl)ation of transcription factor Yin Yang 1 under conditions of DNA damage. *Biochem. Biophys. Res. Commun.* **285**, 27-31.
Ogawa, Y., Sun, B. K. and Lee, J. T. (2008). Intersection of the RNA interference and X-inactivation pathways. *Science* **320**, 1336-1341.
Ohlsson, R., Renkawitz, R. and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520-527.
Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., Klenova, E., Lobanenkov, V. and Ohlsson, R. (2004). Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Mol. Cell. Biol.* **24**, 3497-3504.
Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422-433.
Peric-Hupkes, D. and van Steensel, B. (2008). Linking cohesin to gene regulation. *Cell* **132**, 925-928.
Pugacheva, E. M., Tiwari, V. K., Abdullaev, Z., Vostrov, A. A., Flanagan, P. T., Quitschke, W. W., Loukinov, D. I., Ohlsson, R. and Lobanenkov, V. V. (2005). Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. *Hum. Mol. Genet.* **14**, 953-965.
Reale, A., Matteis, G. D., Galleazzi, G., Zampieri, M. and Caiafa, P. (2005). Modulation of DNMT1 activity by ADP-ribose polymers. *Oncogene* **24**, 13-19.
Regha, K., Sloane, M. A., Huang, R., Pauler, F. M., Warczok, K. E., Melikant, B., Radolf, M., Martens, J. H., Schotta, G., Jenuwein, T. et al. (2007). Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. *Mol. Cell* **27**, 353-366.
Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., Aebersold, R., Ranish, J. A. and Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. USA* **105**, 8309-8314.
Schreiber, V., Dantzer, F., Ame, J. C. and de Murcia, G. (2006). Poly(ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell. Biol.* **7**, 517-528.
Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348-1354.
Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N. and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20**, 2349-2354.
Stedman, W., Kang, H., Lin, S., Kissil, J. L., Bartolomei, M. S. and Lieberman, P. M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J.* **27**, 654-666.
Torrano, V., Navascues, J., Docquier, F., Zhang, R., Burke, L. J., Chernukhin, I., Farrar, D., Leon, J., Berciano, M. T., Renkawitz, R. et al. (2006). Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. *J. Cell Sci.* **119**, 1746-1759.
Uhlmann, F. (2008). Molecular biology: cohesin branches out. *Nature* **451**, 777-778.
Wallace, J. A. and Felsenfeld, G. (2007). We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* **17**, 400-407.
Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T. et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796-801.
Wilkinson, F. H., Park, K. and Atchison, M. L. (2006). Polycomb recruitment to DNA *in vivo* by the YY1 REPO domain. *Proc. Natl. Acad. Sci. USA* **103**, 19296-19301.
Williams, A. and Flavell, R. A. (2008). The role of CTCF in regulating nuclear organization. *J. Exp. Med.* **205**, 747-750.
Xie, X., Mikkelsen, T. S., Gnirke, A., Lindblad-Toh, K., Kellis, M. and Lander, E. S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc. Natl. Acad. Sci. USA* **104**, 7145-7150.
Xu, N., Donohoe, M. E., Silva, S. S. and Lee, J. T. (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat. Genet.* **39**, 1390-1396.
Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C. et al. (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat. Genet.* **36**, 1105-1110.
Yusufzai, T. M. and Felsenfeld, G. (2004). The 5′-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc. Natl. Acad. Sci. USA* **101**, 8620-8624.
Yusufzai, T. M., Tagami, H., Nakatani, Y. and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* **13**, 291-298.
Zhang, R., Burke, L. J., Rasko, J. E., Lobanenkov, V. and Renkawitz, R. (2004). Dynamic association of the mammalian insulator protein CTCF with centrosomes and the midbody. *Exp. Cell. Res.* **294**, 86-93.
Zhao, H. and Dean, A. (2004). An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res.* **32**, 4903-4914.
Zhao, Z., Tavoosidana, G., Sjolinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K. S., Singh, U. et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341-1347.
Zlatanova, J. and Thakar, A. (2008). H2A.Z: View from the top. *Structure* **16**, 166-179.
Zlatanova, J. and Caiafa, P. (2009). CCCTC-binding factor: to loop or to bridge. *Cell. Mol. Life Sci.* (in press).