Published in final edited form as:

Curr Opin Genet Dev. 2007 October ; 17(5): 400–407.

WE GATHER TOGETHER: INSULATORS AND GENOME ORGANIZATION

Julie A. Wallace and Gary Felsenfeld

Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892-0540

Abstract

When placed between an enhancer and promoter, certain DNA sequence elements inhibit enhancer-stimulated gene expression. The best characterized of these enhancer blocking insulators, *gypsy* in *Drosophila* and the CTCF-binding element in vertebrates and flies, stabilize contacts between distant genomic regulatory sites leading to the formation of loop domains. Current results show that CTCF mediates long-range contacts in the mouse β -globin locus and at the *Igf2/H19* imprinted locus. Recently described active chromatin hubs and transcription factories also involve long-range interactions; it is likely that CTCF interferes with their formation when acting as an insulator. The properties of CTCF, and its newly described genomic distribution, suggest that it may play an important role in large scale nuclear architecture, perhaps mediated by the co-factors with which it interacts *in vivo*.

Introduction

Insulators are DNA elements that were first identified based on their ability to protect a gene from outside influences, which might otherwise lead either to inappropriate activation or silencing of the gene. Insulators have been divided into two classes: enhancer blocking (EB) insulators, which prevent distal enhancers from activating a promoter when placed between an enhancer and promoter, and barrier insulators, which block heterochromatinization and consequent silencing of a gene. In both cases, studies have shown that insulators make use of many of the known mechanisms of epigenetic regulation and genome organization. Recent results suggest that insulating activities, though exploited by the cell for important biological purposes, may be incidental to other properties of equal importance to cell function. Particularly in the case of enhancer blocking, the subject of this review, insulation may be a corollary of a wider role in organizing large-scale structures within the nucleus.

Enhancer blocking insulators are best exemplified by the *gypsy* element in *Drosophila* [1] and the CTCF-binding sites identified as insulators initially in vertebrates [2] and later in *Drosophila* [3]. In both cases, placing the insulator element between a variety of enhancers and promoters interferes with transcriptional activation. The ability of CTCF to confer EB insulation was discovered during studies of a DNA insulator element, 5'HS4, located at the 5' end of the chicken β -globin locus [2]. Interest rose with the discovery that CTCF mediates allele-specific expression at the *Igf2/H19* imprinted locus in mouse and human [4–6] (Figure 1A). This finding established the functional importance of EB insulators; much of this report is devoted to recent studies of CTCF function, the structures and co-factors with which it is associated, and its distribution in the genome.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Long-range contacts within the nucleus

A series of papers analyzing the *gypsy* element have been important in establishing what may be a general model for the way EB insulators organize chromatin structure. *Gypsy* sites are bound, in a sequence-specific manner, by a protein, Suppressor of Hairy-wing (Su(Hw)), which recruits other factors. Multiple *gypsy* sites and their associated proteins cluster together to form 'insulator bodies', with the ultimate effect of organizing the nearby chromatin into loop domains (Figure 1B). Although less well studied, the *Drosophila scs/scs*' insulator elements also stabilize loop formation mediated by a different set of proteins [7,8].

Although the mechanisms may be different, the overall concept of loop domain organization, established in *Drosophila*, also appears to hold true for CTCF insulator sites. Before discussing this in detail, it is useful to review what is known about long-range interactions within the nucleus. Much of our knowledge comes from Chromatin Conformation Capture (3C) and its elaborations, measurements that detect contacts between distant DNA sequences, in combination with DNA and RNA fluorescent in situ hybridization (FISH). Studies over a number of years have revealed that enhancers, locus control elements, and promoters are in contact with one another when genes are being transcribed [9]. Other recent studies [10] have made it clear that most or all active genes are transcribed while associated with RNA polymerase (pol) II transcription factories containing many pol II molecules. A given factory is visited by many genes, with the effect of bringing individual genes that are widely separated on a chromosome, or even situated on different chromosomes, into proximity. The formation of loop domains thus appears to be a common and necessary process associated with transcription in eukaryotes.

This point of view has strongly influenced recent papers analyzing CTCF insulator function. CTCF can interact with itself and with other proteins that may allow it to form clusters, which, as in the case of *gypsy*, would lead to formation of discrete domains [11]. Recent data confirm that CTCF molecules bound to distant sites, even on different chromosomes, can interact with one another *in vivo* [12].

Role of CTCF in Long-range Interactions

Physical interactions among *cis*-regulatory elements have been most elegantly demonstrated at the mouse β -globin locus. In differentiated primitive and definitive erythroid cells, where the *globin* genes are expressed, the distant locus control regions (LCRs) cluster together with the developmentally appropriate promoters to form an active chromatin hub (ACH) [9]. These interactions are absent in embryonic day 12.5 erythroid progenitor cells, which do not express *globin*. There are three CTCF-binding sites upstream of the mouse β -globin locus and one downstream [13,14]. Recent 3C results [15] show that, in erythroid progenitor cells, all the CTCF sites are in contact with one another, maintaining a compact domain structure that is dependent on CTCF expression (Fig. 1C). However, at later developmental stages, neither CTCF nor its downstream binding site is necessary for establishment of an ACH or for proper expression of the *globin* genes.

A more complicated picture emerges from 3C analysis of the mouse imprinted *Igf2/H19* locus. Four CTCF-binding sites are situated at the imprinted control region (ICR) that lies between the *Igf2* gene and downstream endodermal and mesodermal enhancers. These sites are occupied by CTCF on the maternally inherited allele, but the DNA of the paternal copy is methylated at the ICR, preventing CTCF binding and abolishing insulator function (Fig. 1A) [4–6]. Mice in which the two alleles can be distinguished have been used in 3C experiments to study insulator effects on higher order structure [16,17]. Contacts are detected between the ICR and one of two upstream, imprinted DNA methylated regions (DMR), DMR1 and DMR2 (Fig. 1D, panels a and b). These contacts differ on the two alleles, suggesting a way in which allele-specific

interaction between the enhancer and an *Igf2* promoter could occur to give rise to allele-specific *Igf2* expression.

A different model of the *Igf2/H19* system has been proposed based on a similar 3C approach, but with emphasis on interactions between promoters and downstream enhancers [18] (Fig. 1D, panels c and d). The paternal *Igf2* promoters are found to contact the endodermal or the mesodermal enhancer in liver and skeletal muscle cells respectively, consistent with the presence of allele-specific ACHs. On the maternal allele, carrying an active insulator, the hub is disrupted. There is no promoter-enhancer contact, but both *Igf2* promoters and downstream enhancers interact instead with the ICR/insulator element. The insulator thus preempts promoter-enhancer interaction. The behavior of the ICR is independent of its location and of the associated gene: predictable and consistent results are obtained when a copy is placed between the endodermal and mesodermal enhancers, or when it is coupled to a different gene and its associated regulatory element.

These experiments suggest a model in which CTCF-binding sites mark the boundaries of separate domains. In the case of the mouse β -globin locus, the presence of such structures in erythroid precursor cells depends on direct or indirect interactions between CTCF molecules bound to insulator sites at opposite ends of the gene cluster [9]. Long-range contacts, dependent on the presence of CTCF at both sites, also stabilize the interaction between the *Igf2* ICR and one of the alleles of the *Wsb1/Nf1* gene pair on another chromosome [12]. It is less clear what interactions stabilize CTCF/insulator interactions with promoters and enhancers within the *Igf2/H19* locus, although imprinted CTCF sites have been documented upstream of *Igf2* [17] and CTCF-binding sites have been reported in the neighborhood of the human *Igf2* promoters [19]. As will be discussed below, CTCF can recruit many co-factors to its binding sites, and the mechanisms stabilizing long-range interaction could be correspondingly complex. This seems likely, given that the gypsy insulator body in *Drosophila*, which requires multiple proteins for its stabilization, depends also on an RNA component (Fig. 1B); mutations of the Argonaute genes *piwi* or *aubergine*, part of the RNAi pathway, impair insulator activity as well as insulator body morphology [20]. It should be noted that a protein with strong homology to CTCF is also present in *Drosophila*, and that it is localized at Fab-8, a known insulator site important for regulation of the *Abd-B* locus [3].

Mechanisms of EB insulator action

The effect of sequestering an enhancer and promoter in separate domains, however they are formed, is to prevent enhancer-promoter interaction. As has been pointed out [21,22], this could be a steric effect, or it could result from interference with a processive activating signal (histone modification, pol II transcription) originating at the enhancer. In either case, the mechanism must be directional: insulation only occurs when the insulator lies between promoter and enhancer. Evidence consistent with the processive model comes from an analysis of an episome carrying the human HS2 enhancer and ϵ -globin gene [23,24]. Insertion of the chicken β -globin 5'HS4 insulator element between an enhancer and promoter results in inhibition of transcription, accumulation of RNA pol II 5' of the insulator, and interruption in the spread of histone H3 and H4 acetylation.

The action of insulators is not absolute. Strong enhancers can overcome insulation [24,25] and promoter targeting sequences (PTS) found in *Drosophila* can nullify insulator action and allow a blocked enhancer to function [26]. Perhaps 3C methods can be used to investigate how PTS elements enter into the complex array of interactions and structures involved in insulator activity. We point out that although this review deals only with EB insulators, barrier insulator elements may also establish loop domains as part of their function (see [27]).

Identifying CTCF-binding sites in the genome

Originally identified as a transcriptional repressor at the *myc* locus, CTCF has since been characterized as a transcriptional activator, enhancer-blocker, boundary definer and, potentially, genome organizer. It remains to be determined, therefore, where CTCF functions within the genome and how it carries out its various activities at specific sites. The ability of CTCF to act with a diversity of functions has been attributed to its structure. CTCF contains 11 zinc-finger domains with which it binds DNA in a sequence-specific manner. Two recent studies have undertaken different approaches to identify CTCF-binding sites throughout the genome and to tighten a consensus sequence for CTCF binding. Ren and colleagues employed ChIP-Chip, immunoprecipitating CTCF-bound DNA and hybridizing to an oligomicroarray, to identify novel CTCF-binding sites in human fibroblasts, whereas Lander and colleagues applied a computational approach towards the identification of conserved regulatory elements in the genome [19,28]. From these studies, ~14,000 CTCF-binding sites were identified, including a number that have been described previously.

Additionally, this genome-wide approach allowed characterization of the distribution of CTCF in the genome. First, CTCF binding is ubiquitous throughout the genome at conserved sites, implying that these sites are functional. Identified sites correlate strongly with regions containing genes, suggesting that CTCF's primary role in the genome is to regulate gene expression, although these CTCF sites are far from promoters. Second, domains depleted of CTCF sites tend to include clusters of related gene families and genes that are transcriptionally coregulated; many of these regions are flanked by a pair of CTCF-binding sites. Domains that are enriched in CTCF sites tend to have multiple alternative promoters [19]. These observations are all consistent with a role for CTCF as an insulator. Finally, both studies provide an excellent resource for pursuing further studies of CTCF-binding sites towards a full catalog of CTCF's functions and mechanisms.

The discovery of a multitude of CTCF-binding sites facilitates generation of a consensus sequence. The identified 20bp consensus binding motif refines that previously described [4]. Base choices at several sites within the motif are ambiguous, and not all of the newly identified sites in the Ren study [19] are a version of this consensus. It is possible that variation within the consensus sequence can generate versatility in CTCF function if there is a difference in the specific zinc fingers required for binding at each variant of the consensus. Consistent with this idea, studies of truncated versions of CTCF's zinc fingers domain have revealed differences in binding among various CTCF sites (reviewed in [29]). It has been suggested, therefore, that CTCF's myriad functions result from specific differences at each site regarding the fingers bound to DNA and those available for protein-protein interactions.

Another recent genome-wide survey of CTCF-binding sites is part of a high-resolution study, using Solexa sequencing technology, of histone modifications in human T cells [30]. In a number of cases, CTCF sites flank both ends of an active gene domain, which displays extensive H3K27 monomethylation. Immediately outside the region bounded by CTCF, however, the H3K27 sites are trimethylated, a modification associated with condensed, inactive chromatin. The CTCF sites thus mark the boundaries of open chromatin domains, as in the case of the chicken *β -globin* locus where the insulator function of CTCF was first noted [2, 31].

Post-translational modifications of CTCF

CTCF's functions may be regulated through its genomic location and/or by the choice of zinc fingers used in DNA binding. In addition, CTCF is post-translationally modified; it can be phosphorylated in its C-terminus and poly(ADP-ribosyl)ated in its N-terminus, and it is likely that more modifications will be identified [32,33]. These post-translational modifications

probably play a key role in regulating CTCF binding and/or in mediating CTCF's diverse functions. Studies using 3-aminobenzamide (ABA), a general inhibitor of poly(ADP-ribose) polymerases, have led to the suggestion that insulator activity requires poly(ADP-ribosyl)ation of CTCF [33]. It should be noted, however, that poly(ADP-ribosyl)ation has been implicated in maintenance of DNA hypomethylation in the genome [34,35]. As CTCF binding is sensitive to methylation status, further work may be needed to reach an unambiguous conclusion.

Recently, poly(ADP-ribosyl)ation of CTCF was implicated in nucleolar targeting of CTCF upon differentiation [36]. In response to various stimuli inducing differentiation of human cell lines, CTCF was observed to relocate to nucleoli from its normal diffuse nuclear distribution [36,37]. Interestingly, the zinc finger domain of CTCF was sufficient for strong nucleolar localization, while the C-terminus conferred only weak localization. Treatment of these cells with ABA led to an exclusion of CTCF from the nucleoli for the full-length CTCF, but not the zinc fingers construct. It seems likely, as the authors suggest, that the translocation of CTCF to the nucleolus is the consequence of protein-protein interactions, perhaps with nucleophosmin/B23 ([11], see below).

CTCF interaction partners

CTCF has a number of identified binding partners and the list continues to grow rapidly (Table 1). As these partners have been implicated in a range of activities, the variety of CTCF functions, including insulation and genome organization, is likely to reflect the diversity of these cofactors. CTCF interacts with DNA-binding proteins and transcription factors, proteins that interact with histones as well as histones themselves, and other regulatory proteins (Table 1). While several of these proteins have been demonstrated to bind the zinc finger domain of CTCF (Sin3, YB-1, Chd8), Kaiso binds the C-terminus and Yy1 has its highest affinity for the N-terminus (Table 1). It should be noted that while interaction with CTCF has been demonstrated for these proteins, only CHD8, nucleophosmin and RNA pol II have been ChIPed at CTCF-dependent insulators, as indicated in Table 1. It remains to be determined, therefore, whether a number of these identified interactions have functional consequences *in vivo*.

It has been suggested that the interaction between CTCF and nucleophosmin is connected with its insulator function [11]. Nucleophosmin was detected at sites of CTCF binding in the chicken β -globin locus, and a multicopy transgene containing the 5'HS4 insulator was localized by FISH to the periphery of the nucleolus in a CTCF-dependent manner. These results led to a model of insulator action in which CTCF-nucleophosmin interactions result in the tethering of the insulator to a nuclear structure, thus prohibiting enhancer-promoter communication. It is not yet known whether CTCF-nucleophosmin interactions occur at other CTCF-binding sites and whether tethering of the insulator to a nuclear structure is sufficient to confer insulator function.

Another partner of CTCF, CHD8, has been implicated directly in insulator function [38]. CHD8, a member of the chromodomain helicase family, interacts with CTCF *in vitro* and is found at several known CTCF-binding sites. Knockdown of CHD8 by RNAi disrupts enhancer blocking in a reporter assay and at the endogenous *H19* insulator, without displacing CTCF from the ICR. CHD8 is known to form complexes with histone modifying enzymes [39], and many members of the CHD family interact both with modifying enzymes and ATP-dependent chromatin remodeling complexes [40]. It thus seems plausible that CHD8 may bring similar activities to CTCF-bound insulators. Enhancer blocking activity could be aided if CHD8 complexes interfered with processive activating signals from an upstream enhancer, or if these complexes helped in recruitment of CTCF sites to nuclear substructures.

CTCF has also been shown to interact *in vitro* with the large subunit of RNA pol II, both RNA pol II α and II β [41]. These two different forms of RNA pol II vary in their phosphorylated state

and in the complexes with which they tend to interact; pol IIa is hypophosphorylated and is part of the initiation complex, whereas pol IIo is hyperphosphorylated and associates with the elongating complexes. CTCF *in vivo* primarily interacts with pol IIa, the form associated with the initiation complex. Studies at the chicken HS4 β -globin insulator revealed an interaction between CTCF and RNA pol II specifically in proliferating cells, but not in differentiated cells [41], and pol II appears to accumulate at the HS4 insulator when it is downstream of an enhancer (see above) [23]. This suggests that, rather than actively recruiting RNA pol II, CTCF could, in these cases, contact RNA pol II as a consequence of blocking the spread of RNA pol II or its transfer from enhancer to promoter.

Conclusions

Many, perhaps most, transcription-related events within the nucleus take place at clusters of regulatory molecules that promote long-range interactions between genes and their regulatory sites. The two best characterized insulator proteins, Su(Hw) and CTCF, are implicated in stabilizing long-range interactions, but these clusters may be independent of the transcription-related centers. CTCF, in particular, can mediate long-range interactions at the β -globin locus at developmental stages where ACHs do not form. It has been suggested that this activity helps to organize the β -globin locus and other gene clusters into domains amenable to shared regulation or to exclusion of outside influences [15]. This is consistent with the observation that many gene domains in the human genome are flanked by CTCF sites. Insulation is an essential function, but it may reflect an even more general role in genome organization.

Long-range interactions involving CTCF can occur between distant pairs of sites that bind CTCF [12,15]; these may be stabilized simply by CTCF dimerization [11]. CTCF may also be able to form loops by interacting with other target proteins at distant sites. These may be mediated by some of the CTCF-associated factors we have discussed. Whether enhancer, promoter and CTCF are all gathered at the same site or at distinct sites is unknown [18]. The implications of loop formation for insulation mechanisms have been examined elsewhere [22]. Problems remain with all of the obvious mechanisms, which suggest that just as there appears to be more than one kind of CTCF-mediated interaction, there may be more than one way to make an insulator.

Acknowledgements

We thank our colleagues in the Felsenfeld lab for their comments on this manuscript. This work was supported by the Intramural Research Program, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

References

1. Capelson M, Corces VG. The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol Cell* 2005;20:105–116. [PubMed: 16209949]
2. Chung JH, Whiteley M, Felsenfeld G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* 1993;74:505–514. [PubMed: 8348617]
3. Moon H, Filippova G, Loukinov D, Pugacheva E, Chen Q, Smith ST, Munhall A, Grewe B, Bartkuhn M, Arnold R, et al. CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep* 2005;6:165–170. [PubMed: 15678159]
4. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 2000;405:482–485. [PubMed: 10839546]
5. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 2000;405:486–489. [PubMed: 10839547]

6. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol* 2000;10:853–856. [PubMed: 10899010]
7. Blanton J, Gaszner M, Schedl P. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev* 2003;17:664–675. [PubMed: 12629048]
8. Hart CM, Zhao K, Laemmli UK. The scs' boundary element: characterization of boundary element-associated factors. *Mol Cell Biol* 1997;17:999–1009. [PubMed: 9001253]
9. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 2003;35:190–194. [PubMed: 14517543]
10. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 2004;36:1065–1071. [PubMed: 15361872]
11. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* 2004;13:291–298. [PubMed: 14759373]
12. Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, Cherry AM, Hoffman AR. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* 2006;312:269–272. [PubMed: 16614224]
13. Farrell CM, West AG, Felsenfeld G. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol* 2002;22:3820–3831. [PubMed: 11997516]
14. Bulger M, van Doorninck JH, Saitoh N, Telling A, Farrell C, Bender MA, Felsenfeld G, Axel R, Groudine M. Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. *Proc Natl Acad Sci U S A* 1999;96:5129–5134. [PubMed: 10220430]
15. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 2006;20:2349–2354. [PubMed: 16951251] ●● A careful study by 3C methods of the role of CTCF in stabilization of long-range interactions within the mouse *β-globin* locus. At an early stage of erythroid cell development, before the onset of *β-globin* gene expression, CTCF-binding sites flanking the locus at either end are in physical contact, in a CTCF-dependent manner. However, these contacts are not essential for later establishment of an ACH
16. Murrell A, Heeson S, Reik W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet* 2004;36:889–893. [PubMed: 15273689]
17. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A* 2006;103:10684–10689. [PubMed: 16815976] ●● A 3C analysis of long-range interactions at the mouse *Igf2/H19* locus taking advantage of its imprinted status. This allows an internally controlled comparison between contacts on the maternally and paternally inherited chromosomes, which reveals major CTCF-dependent differences between them. A model of these interactions is presented; an earlier and less detailed analysis is given in [16] (see also Fig. 1D)
18. Yoon YS, Jeong S, Rong Q, Park KY, Chung JH, Pfeifer K. Analysis of the H19ICR insulator. *Mol Cell Biol* 2007;27:3499–3510. [PubMed: 17339341] ●● A study by 3C methods of the same *Igf2/H19* locus analyzed in [17]. Comparison between alleles shows that contacts between the enhancer and active *Igf2* promoters on the paternal allele are replaced on the maternal allele by contacts between the ICR (occupied by CTCF) and both the enhancer and promoter (see Fig. 1D). Experiments in which copies of the ICR, or reporters containing the ICR, are moved to other genomic sites give results consistent with this picture. There is some disagreement between this paper and [17] concerning enhancer contacts, stemming from a difference in the identification of the enhancer element.
19. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 2007;128:1231–1245. [PubMed: 17382889] ●● Results of tiling microarray analysis of the

genome-wide distribution of bound CTCF in human primary fibroblasts, showing that the distribution of CTCF sites is consistent with widespread activity of CTCF as an insulator, and suggesting that they are well situated to organize the genome into separate clusters of related genes

20. Lei EP, Corces VG. RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat Genet* 2006;38:936–941. [PubMed: 16862159]●●The most recent in a series of papers from this laboratory investigating the genetic and molecular basis of *gypsy* insulator action, and its relationship to clustered loop formation. Here the role of components of the RNAi pathway is demonstrated: the products of *aubergine* and *piwi* stabilize the structure and increase insulating activity (see Fig. 1B)
21. Bell AC, Felsenfeld G. Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev* 1999;9:191–198. [PubMed: 10322134]
22. Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 2006;7:703–713. [PubMed: 16909129]
23. Zhao H, Dean A. An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res* 2004;32:4903–4919. [PubMed: 15371553]
24. Zhao H, Kim A, Song SH, Dean A. Enhancer blocking by chicken beta-globin 5′-HS4: role of enhancer strength and insulator nucleosome depletion. *J Biol Chem* 2006;281:30573–30580. [PubMed: 16877759]●A CTCF-dependent insulator blocks the spread of positive histone modifications associated with the presence of an upstream enhancer, evidence that the insulator blocks a processive activating signal from reaching the promoter
25. Scott KC, Taubman AD, Geyer PK. Enhancer blocking by the *Drosophila* gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics* 1999;153:787–798. [PubMed: 10511558]
26. Zhou J, Levine M. A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the *Drosophila* embryo. *Cell* 1999;99:567–575. [PubMed: 10612393]
27. Noma K, Cam HP, Maraia RJ, Grewal SI. A role for TFIIC transcription factor complex in genome organization. *Cell* 2006;125:859–872. [PubMed: 16751097]
28. Xie X, Mikkelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc Natl Acad Sci U S A* 2007;104:7145–7150. [PubMed: 17442748]●A genome-wide search for conserved non-coding elements reveals about 15,000 conserved sites that correspond to a CTCF-binding consensus. Nearby genes separated by these sites show markedly reduced correlation in gene expression, consistent with a role for CTCF in partitioning the genome
29. Ohlsson R, Renkawitz R, Lobanikov V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet* 2001;17:520–527. [PubMed: 11525835]
30. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell* 2007;129:823–837. [PubMed: 17512414]●●A remarkable survey, making use of Solexa 1G sequencing technology, of genome-wide patterns of histone H3 and H4 methylation in primary human T cells, with correlation to patterns of gene expression and RNA pol II binding. This study also determines the distribution of CTCF-binding sites and relates them to local histone modifications and the presence of the histone variant H2A.Z
31. Saitoh N, Bell AC, Recillas-Targa F, West AG, Simpson M, Pikaart M, Felsenfeld G. Structural and functional conservation at the boundaries of the chicken beta-globin domain. *Embo J* 2000;19:2315–2322. [PubMed: 10811622]
32. Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, Neiman PE, Lobanikov VV. 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* 1993;13:7612–7624. [PubMed: 8246978]
33. Yu W, Ginjala V, Pant V, Chernukhin I, Whitehead J, Docquier F, Farrar D, Tavoosidana G, Mukhopadhyay R, Kanduri C, et al. Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat Genet* 2004;36:1105–1110. [PubMed: 15361875]
34. Zardo G, D’Erme M, Reale A, Strom R, Perilli M, Caiafa P. Does poly(ADP-ribosyl)ation regulate the DNA methylation pattern? *Biochemistry* 1997;36:7937–7943. [PubMed: 9201939]

35. Zardo G, Caiafa P. The unmethylated state of CpG islands in mouse fibroblasts depends on the poly (ADP-ribosyl)ation process. *J Biol Chem* 1998;273:16517–16520. [PubMed: 9632720]
36. Torrano V, Navascues J, Docquier F, Zhang R, Burke LJ, Chernukhin I, Farrar D, Leon J, Berciano MT, Renkawitz R, et al. Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. *J Cell Sci* 2006;119:1746–1759. [PubMed: 16595548]
37. Zhang R, Burke LJ, Rasko JE, Lobanenko V, Renkawitz R. Dynamic association of the mammalian insulator protein CTCF with centrosomes and the midbody. *Exp Cell Res* 2004;294:86–93. [PubMed: 14980504]
38. Ishihara K, Oshimura M, Nakao M. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* 2006;23:733–742. [PubMed: 16949368]●●Identification of an interaction between CHD8 and CTCF *in vitro* and *in vivo*. CHD8 is located at several known CTCF-binding sites and is required for insulation at the *H19/Igf2* ICR. CHD family members associate with complexes containing chromatin remodeling and histone modification activities, which may participate in establishing insulation
39. Dou Y, Milne TA, Tackett AJ, Smith ER, Fukuda A, Wysocka J, Allis CD, Chait BT, Hess JL, Roeder RG. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* 2005;121:873–885. [PubMed: 15960975]
40. Marfella CG, Imbalzano AN. The Chd family of chromatin remodelers. *Mutat Res* 2007;618:30–40. [PubMed: 17350655]
41. Chernukhin I, Shamsuddin S, Kang SY, Bergstrom R, Kwon YW, Yu W, Whitehead J, Mukhopadhyay R, Docquier F, Farrar D, et al. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol Cell Biol* 2007;27:1631–1648. [PubMed: 17210645]
42. Chernukhin IV, Shamsuddin S, Robinson AF, Carne AF, Paul A, El-Kady AI, Lobanenko VV, Klenova EM. 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* 2000;275:29915–29921. [PubMed: 10906122]
43. Klenova E, Scott AC, Roberts J, Shamsuddin S, Lovejoy EA, Bergmann S, Bubbs VJ, Royer HD, Quinn JP. YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. *J Neurosci* 2004;24:5966–5973. [PubMed: 15229244]
44. Defossez PA, Kelly KF, Filion GJ, Perez-Torrado R, Magdinier F, Menoni H, Nordgaard CL, Daniel JM, Gilson E. The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J Biol Chem* 2005;280:43017–43023. [PubMed: 16230345]
45. Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT. Identification of a Ctf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell* 2007;25:43–56. [PubMed: 17218270]
46. Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, Arnold R, Schultheiss H, Brehm A, Kouzarides T, Lobanenko V, et al. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res* 2000;28:1707–1713. [PubMed: 10734189]

Figure 1. Loop domains and enhancer blocking insulators

A. Action of CTCF at the mouse *Igf2/H19* locus. On the maternally transmitted allele, CTCF binds to sites in the imprinted control region (ICR) and prevents downstream enhancers from activating *Igf2* expression. The ICR of the paternal allele is methylated, CTCF does not bind, and the enhancer is no longer blocked [4–6].

B. Loop formation stabilized by proteins and RNA bound to *gypsy* sites in *Drosophila* [20] (Figure taken from [20]; see [22] for review of a more complete version of this complex). For clarity, one set of interactions, leading to the formation of a single loop, is shown. Clusters of such sites form insulator bodies *in vivo*. (a) The proteins recruited to the site include the DNA-binding protein Su(Hw), CP190, and Mod(mdg4)2.2, as well as Topors, a ubiquitin ligase, which associates with the nuclear lamina and the insulator complex and is required for insulator function. (b,c) Loop formation and insulator activity also involve RNA and are dependent on members of the RNAi processing pathway. (d) Loop formation is interfered with by the RNA-binding protein, Rm62. The protein CP190 is additionally involved in insulator function at sites that do not contain Su(Hw).

C. Contacts between distant CTCF-binding sites (red circles), which are also nuclease hypersensitive sites (HS), in embryonic day 12.5 mouse erythroid progenitor cells [9,15]. HS-62 is ~ 62kb upstream, and HS5 ~ 30kb upstream, of the *globin* gene *Ey*. 3'HS1 is downstream of the *globin* gene array. Other hypersensitive sites (yellow) mark locus control regions; not all such sites are shown. The *globin* genes (blue squares) are not expressed in these cells. Mutating the CTCF site at 3'HS1 destroys the interactions between it and the other two CTCF sites. At a later stage of erythroid development additional contacts are made between locus control region and other regulatory sites to form the ACH; this structure is independent of CTCF binding [15].

D. Loop domain models for the action of the ICR at the mouse *Igf2/H19* locus, showing contacts involving the ICR, detected by 3C and other methods, that differ on the maternal and paternal alleles [15,18]. It is proposed that the conformation of the paternal allele allows contacts between *Igf2* promoters and the enhancer that are blocked on the maternal allele. (a,b) The ICR on the maternal allele (a) makes contact with an imprinted site (DMR1) located upstream of the *Igf2* promoters. On the paternal allele (b) the ICR contacts another such site, DMR2, downstream of *Igf2*. This should make the *Igf2* promoters accessible to the enhancers located downstream of *H19*. For a more elaborate model based on additional data, see [16]. (c,d) A second set of 3C experiments [18] shows that on the maternal allele (c) one of the downstream enhancers (mesodermal or endodermal specific, depending on cell type) contacts the ICR, effectively inhibiting contact between enhancer and *Igf2* promoter. Other interactions, not shown, occur between the ICR and the *Igf2* promoter. On the paternal allele (d), with CTCF not bound to the ICR, the enhancer makes contact with the appropriate subset of *Igf2* promoters. Differences between the results shown in (a,b) and (c,d) may reflect different choices of anchor and target sites for the 3C analysis, and may not be mutually exclusive. Some of these interactions may be tissue-specific; which interactions are most important for establishing allele-specific expression is not yet clear.

Table 1

Identified interaction partners of CTCF

Protein Name	Protein Function	CTCF Interaction Domain	References
<i>DNA-binding proteins and transcription factors</i>			
YB-1	DNA/RNA-binding factor	Zinc fingers	[42,43]
Kaiso	Zinc finger transcription factor	C-terminus	[44]
Yy1	Zinc finger transcription factor	All, but highest affinity for N-terminus	[45]
<i>Histones and Histone modifying proteins</i>			
Sin3	Transcriptional repressor, associated with histone deacetylases?	Zinc fingers	[46]
CHD8 [*]	Chromodomain helicase family member	Zinc fingers	[38]
Taf-1/Set	Molecular chaperone, associated with inhibition of histone acetyltransferases?	N/A	[11]
H2A/H2A.Z	Nucleosome component	N/A	[11,30]
<i>Other regulatory proteins</i>			
Nucleophosmin [*]	Molecular transport chaperone	N/A	[11]
Poly(ADP-ribose) polymerase	Nuclear enzyme	N/A	[11]
Topoisomerase II	DNA topology enzyme	N/A	[11]
Lamin A/C	Nuclear membrane component	N/A	[11]
Importin $\alpha 3/\alpha 1$	Nuclear import chaperone	N/A	[11]
RNA polymerase II (large subunit) [*]	Catalyzes transcription of DNA	C-terminus	[41]

* these proteins have been identified at CTCF-dependent insulators