Journal of Biology

Minireview

Insulators as mediators of intra- and inter-chromosomal interactions: a common evolutionary theme

Chin-Tong Ong and Victor G Corces

Address: Department of Biology, Emory University, 1510 Clifton Road NE, Atlanta, GA 30322, USA.

Correspondence: Victor G Corces. Email: vcorces@emory.edu

Abstract

Insulator elements mediate intra- and inter-chromosomal interactions. The insulator protein CCCTC-binding factor (CTCF) is important for insulator function in several animals but a report in *BMC Molecular Biology* shows that *Caenorhabditis elegans*, yeast and plants lack CTCF. Alternative proteins may have a similar function in these organisms.

Eukaryotic genomes have developed a variety of strategies for efficiently orchestrating the complex patterns of gene expression required for proper cellular differentiation. Comparative genome analyses suggest that developmental evolution is largely driven by the increase in the complexity of these expression patterns [1]. Consistent with this hypothesis, recent studies indicate that transcription factor-coding genes tend to be under greater positive evolutionary selection compared with other genes [2]. To establish and maintain cell-specific patterns of gene expression, regions of the genome are kept in a silenced state while immediately adjacent regions are transcriptionally active because of the presence of promiscuous enhancer elements that can act over large distances. Insulators were originally described as DNA regulatory elements that ensure the progress of an accurate transcriptional program by keeping in check communication between enhancers and promoters and creating boundaries that prevent inappropriate interactions between adjacent chromatin domains. Accumulating evidence suggests that these properties of insulators arise from their ability to mediate intra- and inter-chromosomal interactions, which result in the formation of chromatin loops through clustering of multiple insulator sites [3]. Depending on the complexity of the genome, the capability to mediate long-range interactions with other protein complexes may allow insulator proteins to carry out a variety of functions in the nucleus [4].

CCCTC-binding factor (CTCF) is the only known insulator protein necessary for establishing patterns of nuclear architecture and transcriptional control in vertebrates [5]. This protein is also found in invertebrates such as *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster* [6]. A recent study by Heger *et al.* in *BMC Molecular Biology* [7] has shown that the gene encoding CTCF is not present in the genomes of several model organisms, including *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana* and *Caenorhabditis elegans*. Because of the widespread presence of insulators and the essential role of CTCF in a wide variety of eukaryotic organisms, this absence of the gene in other organisms raises the possibility that other regulatory mechanisms might have evolved to replace the function of this protein. Here, we provide a brief overview of how insulator proteins work in *Drosophila* and vertebrates, as well as how plants and fungi may have adapted different proteins to accomplish insulator function. We also discuss how insulator proteins such as CTCF may have evolved new functions to handle more complex genomes in animals.

Examples of insulator function

The mechanisms of insulator function are best understood from analyses of the *gypsy* element of *Drosophila*. *Gypsy* insulator sites are bound by the Suppressor of Hairy-wing protein (Su(Hw)), in a sequence-specific manner. This protein in turn recruits other factors, including centrosomal protein 190 kDa (CP190), Modifier of mdg4 (Mod(mdg4)2.2), topoisomerase I-interacting RS protein (dTopors) and RNA, to form clusters of 'insulator bodies' (consisting of these proteins and DNA) with multiple *gypsy* sites [8] (Figure 1a). Recently, other *Drosophila* insulator proteins, dCTCF and Boundary element associated factor (BEAF), have also been shown to recruit CP190 to specific DNA sites [9], suggesting that loop formation through long-range protein interactions mediated by CP190 might be the underlying mechanism for insulator function in *Drosophila*.

The concept of intra- and inter-chromosomal interaction mediated by insulator proteins in *Drosophila* seems to be applicable to the CTCF insulator in vertebrates, despite the involvement of a different set of protein complexes. The mechanism of CTCF function in vertebrates is best illustrated by the mouse imprinted *Igf2-H19* locus [3], where four CTCF-binding sites are located at the imprinted

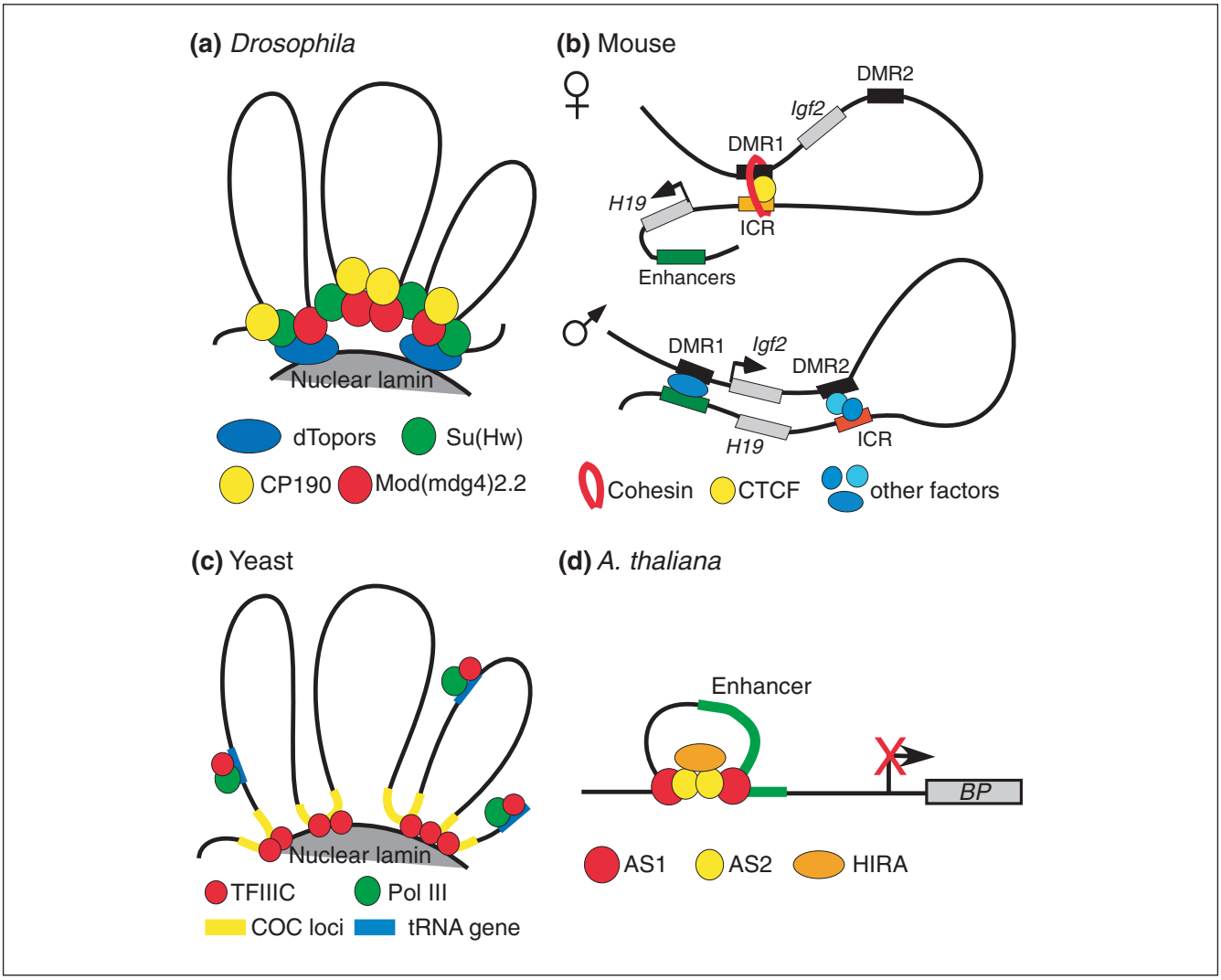

Figure 1

Loop formation through intra- and inter-chromosomal interactions is a common strategy for genome organization and insulation in different organisms. **(a)** In *Drosophila*, the Su(Hw) protein binds to specific DNA elements and recruits the CP190 protein and Mod(mdg4)2.2 proteins. Interaction among these proteins results in the formation of chromatin loops. Mod(mdg4)2.2 attaches the chromatin to the nuclear periphery through its interaction with topoisomerase I-interacting RS protein (dTopors). **(b)** Monoallelic expression at the *Igf2-H19* locus is regulated by binding of CTCF to the imprinted control region (ICR). On the maternal allele, CTCF mediates interactions between ICR and DNA methylated region 1 (DMR1) that also involve joining of the DNA strands by cohesin, insulating *Igf2* from the influence of downstream enhancers. Methylated ICR sequences prevent CTCF from binding to the ICR on the paternal allele, allowing downstream enhancers to switch on *Igf2* transcription. **(c)** In *S. pombe*, TFIIIC binds to RNA polymerase (Pol) III at tRNA genes and acts as a barrier against the spreading of heterochromatin. It is also hypothesized to organize the chromatin into distinct loops by clustering various chromosome-organizing clamp (*COC*) loci to the nuclear periphery. **(d)** In *A. thaliana*, binding of the ASYMMETRIC LEAVES1 (AS1)-AS2 complex at two specific DNA sites flanking the enhancer is required to silence the expression of the *BP* gene. Recruitment of the histone chaperone HIRA is necessary for this process, and it probably acts by facilitating looping of the enhancer element.

control region (ICR) that lies between the *Igf2* gene and its downstream enhancers (Figure 1b). CTCF binds to these sites on the maternally inherited allele but not on the methylated paternal copy. Chromatin conformation capture (3C) experiments revealed distinct long-range chromosomal interactions that are specific to the parent of origin (Figure 1b). On the maternal allele, a CTCF-dependent loop formed by contacts between DNA methylated region 1 (DMR1) and the ICR allows downstream enhancers to turn on the *H19* gene. However, on the paternal allele, contacts between DMR2 and ICR allow downstream enhancers to activate the *Igf2* gene. Given that CP190 protein has been shown to interact with CTCF in *Drosophila*, what proteins could then mediate CTCF-dependent looping

of chromatin in vertebrates? Recent data indicate that cohesin might be required for CTCF insulator function [10]. Cohesin complexes mediate cohesion between sister chromatids by connecting two distinct DNA molecules physically. It is therefore plausible that cohesin can create or stabilize DNA loops during interphase by physically connecting different CTCF-binding sites on the same or different DNA molecules, in a manner similar to CP190 and Mod(mdg4) proteins in *Drosophila*.

If CTCF or functionally similar proteins have a role in establishing patterns of nuclear organization by mediating intra- and inter-chromosomal interactions, how do organisms that lack CTCF homologs accomplish the same goal? In *S. pombe* and *S. cerevisiae*, the transcription factor TFIIIC seems to have this role. In fission yeast, binding of TFIIIC to B-box sequences in the inverted repeat boundary elements can prevent the spreading of heterochromatin from the silenced mating-type loci to neighboring euchromatic regions [11]. Detailed genome-wide analyses reveal that TFIIIC associates with RNA polymerase (Pol) III on all tRNA genes, which are mostly found at pericentromeric heterochromatin domain boundaries. In addition, TFIIIC binds to many sites between divergent promoters in the absence of Pol III and acts as a chromosome-organizing clamp (COC) by tethering distant loci to the nuclear periphery [11] (Figure 1c). Similarly, TFIIIC recruited to tRNA genes in budding yeast can act as both an enhancer-blocking insulator and a heterochromatin barrier by preventing ectopic spreading of Sir protein-mediated silencing [12]. These results uncover a general mechanism of genome organization involving the conserved TFIIIC complex in yeast.

Studies of the process by which *KNOTTED1*-like homeobox (*KNOX*) genes are silenced during organogenesis suggest that *A. thaliana* may also use chromatin looping as a way of regulating gene expression [13]. Stable *KNOX* gene silencing requires the DNA-binding proteins ASYMMETRIC LEAVES1 (AS1) and AS2 and the chromatin-remodeling factor HIRA. AS1 and AS2 form a repressor complex that binds directly to two DNA motif sites that flank the enhancer element of the *KNOX* genes *BREVIPEDICELLUS* (*BP*) and *KNOTTED*-like *Arabidopsis* (*KNAT2*) . Interaction between AS1-AS2 complexes at these two sites is required to repress *BP* expression. These results suggest that AS1-AS2 complexes interact to create a loop in the *KNOX* promoter and, through recruitment of HIRA, to form a repressive chromatin state that blocks enhancer activity during organogenesis (Figure 1d). This regulatory mechanism, which may be conserved among plants with compound leaves, is conceptually similar to the action of an insulator in *Drosophila* and vertebrates.

Recent phylogenetic studies using the zinc-finger protein sets from 35 completely sequenced nematodes [7] has discovered the presence of CTCF-like genes in only three basal nematodes and not in other derived nematodes such as *C. elegans*. This suggests that CTCF might have been lost during nematode evolution, probably as a result of a switch from gene regulatory mechanisms involving distantly acting elements and chromatin insulation to polycistronic transcriptional units [7]. However, the presence of higher-order genome organization in yeast suggests the possibility that other protein complexes may have evolved to replace CTCF functions in *C. elegans*.

Common themes

The underlying theme governing insulator function seems to be the establishment of intra- and inter-chromosomal interactions that bring different sequences in close proximity within the nucleus to accomplish a variety of outcomes [4]. Different eukaryotes may have evolved unique machineries to achieve this. It is also clear that insulator proteins such as CTCF may have acquired additional functions with increased complexity of the genome (reviewed in [4]). In yeast (*S. cerevisiae*), which has a haploid genome size of 13 megabases, the primary insulator function of TFIIIC seems to be the demarcation of chromatin into distinct domains for blockage of heterochromatin silencing. In *A. thaliana*, in which genes are only infrequently interrupted by repetitive elements outside the centromeric regions, AS1-AS2 complexes may mainly act to regulate enhancer-promoter interactions. Long-range interactions mediated by insulator proteins have wider functional implications for *Drosophila* and mammals. In *Drosophila*, different insulators have diverse DNA occupancy patterns with respect to gene features, suggesting that the various insulator functions have diversified by using different insulator DNA-binding proteins with a common interacting partner [9]. Interestingly, vertebrate cells, which contain a larger genome that requires more complex forms of regulation, seem to require CTCF to have a wider set of regulatory roles. These include transcriptional regulation of gene expression at the *major histocompatibility complex class II*, *β-globin* and *interferon-γ* loci, *V(D)J* recombination at the immunoglobulin-encoding *Igh* and *Igk* loci, mono-allelic expression of imprinted genes and X-chromosome inactivation [4]. The ability to have such varied roles must rely on context-dependent interactions with a variety of partners. Their identification remains one of the future challenges for the field.

Acknowledgements

Work in the authors' laboratory is supported by Public Health Service Award GM35463 from the National Institutes of Health.

References

1. Shubin N, Tabin C, Carroll S: **Deep homology and the origins of evolutionary novelty.** *Nature* 2009, **457:**818-823.
2. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM: **A census of human transcription factors: function, expression and evolution.** *Nat Rev Genet* 2009, **10:**252-263.

3. Wallace JA, Felsenfeld G: **We gather together: insulators and genome organization.** *Curr Opin Genet Dev* 2007, **17:** 400-407.
4. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137:**1194-1211.
5. Hore TA, Deakin JE, Marshall Graves JA: **The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes.** *PLoS Genet* 2008, **4:**e1000169.
6. Gray CE, Coates CJ: **Cloning and characterization of cDNAs encoding putative CTCFs inthe mosquitoes, *Aedes aegypti* and *Anopheles gambiae*.** *BMC Mol Biol* 2005, **6:**16.
7. Heger P, Marin B, Schierenberg E: **Loss of the insulator protein CTCF during nematode evolution.** *BMC Mol Biol* 2009, **10:**84.
8. Bushey AM, Dorman ER, Corces VG: **Chromatin insulators: regulatory mechanisms and epigenetic inheritance.** *Mol Cell* 2008, **32:**1-9.
9. Bushey AM, Ramos E, Corces VG: **Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions.** *Genes Dev* 2009, **23:**1338-1350.
10. Wendt KS, Peters JM: **How cohesin and CTCF cooperate in regulating gene expression.** *Chromosome Res* 2009, **17:** 201-214.
11. Noma K, Cam HP, Maraia RJ, Grewal SI: **A role for TFIIIC transcription factor complex in genome organization.** *Cell* 2006, **125:**859-872.
12. Simms TA, Dugas SL, Gremillion JC, Ibos ME, Dandurand MN, Toliver TT, Edwards DJ, Donze D: **TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in *Saccharomyces cerevisiae*.** *Eukaryot Cell* 2008, **7:**2078-2086.
13. Guo M, Thomas J, Collins G, Timmermans MC: **Direct repression of KNOX loci by the ASYMMETRIC LEAVES1 complex of *Arabidopsis*.** *Plant Cell* 2008, **20:**48-58.

Published: 27 August 2009
doi:10.1186/jbiol65