

UDC 577.214.6

# Transcription factor CTCF and mammalian genome organization

E. S. Kotova, S. B. Akopov, E. D. Sverdlov, L. G. Nikolaev

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences  
16/10, Miklukho-Maklaya Str., Moscow, Russian Federation, 117997

lev@ibch.ru

---

*The CTCF transcription factor is thought to be one of the main participants in various gene regulatory networks including transcription activation and repression, formation of independently functioning chromatin domains, regulation of imprinting etc. Sequencing of human and other genomes opened up a possibility to ascertain the genomic distribution of CTCF binding sites and to identify CTCF-dependent cis-regulatory elements, including insulators. In the review, we summarized recent data on CTCF functioning within a framework of the chromatin loop domain hypothesis of large-scale regulation of the genome activity. Its fundamental properties allow CTCF to serve as a transcription factor, an insulator protein and a dispersed genome-wide demarcation tool able to recruit various factors that emerge in response to diverse external and internal signals, and thus to exert its signal-specific function(s).*

*Keywords: transcription factor CTCF, chromatin, transcription regulation.*

---

**Introduction.** CCCTC binding factor (CTCF) is an evolutionally conserved transcription factor of vertebrates. It binds to different functional elements of the genome and performs various regulatory functions (see recent reviews [1–3]).

The transcription factor CTCF was first identified as a protein able to recognize the sequence of three direct CCCTC repeats in the regulatory region of chicken *MYC* gene [4, 5]. Concurrently, the NeP1 protein binding to F1 element of the chicken lysozyme silencer was described [6]. NeP1 and CTCF were later found to be the same protein [7].

The primary structure of CTCF is highly conserved; CTCF orthologs were revealed in different vertebrates: humans, dogs, mice, chickens, frogs, etc. [8–10]. The amino acid (a. a.) sequences of human and chicken CTCF proteins are 93 % identical [11].

Human CTCF polypeptide chain consists of 727 a. a. The central DNA-binding domain contains eleven zinc fingers, which are flanked by lysine- and arginine-rich

positively charged N- and C-terminal domains of 267 and 150 a. a. [12]. Following DNA-binding domain there is a glycine-rich motif typical for both ATP- and GTP-binding proteins. The nuclear localization signal is located closer to the C-terminus of the polypeptide, the phosphorylation sites are next to it [9, 11].

The variety of functions of CTCF protein presumes that there is the process of regulation of the CTCF activity, performed, inter alia, via its posttranslational modifications. Noteworthy are such modifications as the phosphorylation by casein kinase 2 [13], ADP-ribosylation [14], and sumoylation [15–17].

In the course of evolution, CTCF appears in bilateral metazoa and is absent in plants and protozoa [18]. CTCF is constitutively expressed and required for the functioning of many vertebrate cell types in the multicellular organism. It is likely that CTCF is not required for the growth of mammalian cells in culture [2].

However, the significance of CTCF protein for the development of vertebrates is evidenced by the fact that murine embryos, homozygous for the knocked out *CTCF* gene, died prior to the implantation [19, 20]. The knock-

out of *CTCF* gene in murine oocytes prevented the normal development of the blastocyst after the fertilization [21].

**The functional properties of CTCF.** *CTCF binding site.* The development of such technologies as ChIP-chip (chromatin immunoprecipitation on chip) and ChIP-seq (chromatin immunoprecipitation-sequencing) [22–24] allowed obtaining a number of consensus DNA sequences that CTCF binds preferentially [22, 23, 25]. It was revealed that both the amino acid sequence of CTCF protein and the nucleotide sequence of its binding site are highly conserved in different kinds of vertebrates.

Later the ChIP-seq procedure was expanded with the exonuclease treatment (ChIP-exo) [26], which secured more precise localization of CTCF binding sites in the genome. It was demonstrated that CTCF binding site may be subdivided into four blocks, each having its own consensus sequence. About half of CTCF binding sites contained only two central blocks. Other CTCF binding sites contained either all the four blocks or the combination of two or three blocks, or just one block. Further studies led to the notion of the CTCF core binding motif [27], which is a base of the majority of binding sites (Fig. 1).

CTCF binding sequences always include a small number of sequences without any detectable consensus [26, 27]. These sequences are assumed to bind CTCF via intermediate proteins rather than directly [27]. Experimentally detected CTCF binding sequences are collected in the CTCF Binding Sites Database [28].

*The interaction of CTCF and DNA.* The electrophoretic mobility shift assay (EMSA) was used to study the interaction of CTCF zinc finger domain fragments and some known CTCF binding sequences, including the site from chicken beta-globin locus [8], sites from murine *Igf2/H19* locus [29, 30] and site APBbeta, located in the promoter of human beta-amyloid precursor [31].

Since zinc fingers are a characteristic DNA-binding structure of many nuclear proteins [32], it was assumed that the ZnF-containing domain is responsible for the CTCF-DNA interaction. It was demonstrated that four zinc fingers (from 4 to 7) are the minimal set required for the specific interaction of CTCF and its binding site *in vitro*. Further reduction in the number of zinc fingers resulted in a sharp decrease in the DNA-binding ability of the CTCF fragment [30].

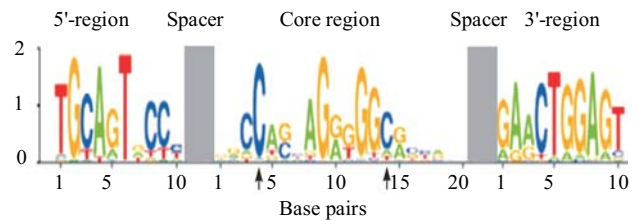


Fig. 1. The general nucleotide structure of the CTCF binding site [27]. The arrows show the potential sites of CpG methylation

CTCF may use different combinations of zinc fingers for binding to DNA. In particular, the binding to the regulatory site of chicken *MYC* gene is effected by zinc fingers from 2 to 7, while the site located close to P2 promoter of human gene *MYC* is bound by CTCF via zinc fingers from 3 to 11 [11]. The binding of chicken lysozyme silencer fragment F1 *in vitro* requires zinc fingers from 5 to 8 [7].

Later these data were confirmed and expanded using mass sequencing [27]. It was demonstrated that the zinc finger eight is not required for the specific recognition of the binding site by CTCF and rather stabilizes the CTCF-DNA complex via unspecific interaction. Zinc fingers 9–11 are responsible for the interaction with a small region of CTCF binding sequence separated by the spacer from the core unit. This fragment, called a U-element (upstream element), corresponds to the block 1 identified in [26]. It is noteworthy that the removal of zinc fingers 8–11 leads to the complete lack of the CTCF interaction with the binding site containing the U-element [27]. It is likely that when a core sequence is far from the consensus, a DNA-protein complex is stable only in case of additional interaction of CTCF and the U-element. Zinc finger 3 is involved in the interaction with DNA in the absence of the U-element, while zinc fingers 1 and 2 are likely to not participate in the specific interaction with DNA, but are responsible for the overall stabilization of the DNA-protein complex.

The hypothetical mechanisms of the loop formation (Fig. 2), via which CTCF may participate in the formation of the chromatin domains and function as a protein component of insulators are suggested [33, 34].

*The impact of DNA CpG methylation on the CTCF binding.* It was demonstrated that CpG methylation of the CTCF binding sites leads to the suppression of their binding to CTCF protein *in vitro* [30, 35–37]. Additio-

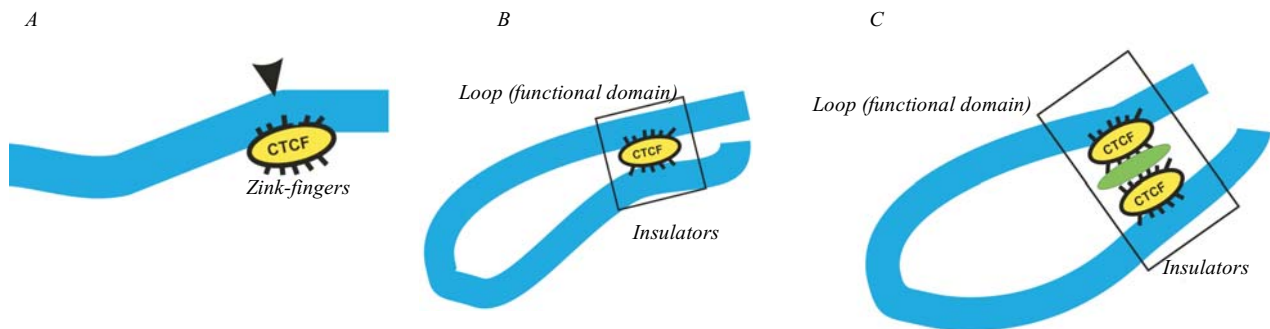


Fig. 2. The hypothetical model of the CTCF-mediated loop formation: *A* – the CTCF binding to its site (shown with an arrow) by means of a subset of its zinc fingers; *B* – using a subset of zinc fingers the protein binds to its site, the remaining zinc fingers are available for the loop formation (DNA bends while binding CTCF which allows the available zinc fingers to close the loop [33]); *C* – the formation of a loop by means of two CTCF molecules binding with two sites, located at the borders of the functional domain (the formation is possible with and without the intermediate protein (a green oval); free zinc fingers may participate in the protein-protein interactions [20, 34]; the regions within the box correspond to the insulators or border elements of the domain)

nally, CTCF does not bind to the methylated ICR region of the paternal allele locus *in vivo* [38, 39], and CTCF binding to ICR of *Igf2/H19* locus of the maternal chromosome hinders its methylation in course of the organism development [39].

Wang et al. [40] investigated the *in vivo* differential methylation of CTCF binding sequences in 19 types of cultured cells and human tissues using ChIP-seq and mass bisulfite sequencing. 36 % of CTCF binding sites were occupied in all 19 types of cells. The inverse relation between the degree of methylation and the ability of binding CTCF was demonstrated for 67 % of differentially methylated CTCF binding sites. Therefore, DNA methylation is an important factor defining CTCF protein binding to a certain nucleotide sequences. It was also demonstrated that about 29 % of CTCF binding sequences in the genome contain CpG at least in one of two positions – 1 and 11, which corresponds to the positions 4 and 14 of the core sequence of the CTCF binding site (Fig. 1, see [27]) and positions –5 and 5 [26].

**Interaction of CTCF with proteins.** During the immunoprecipitation from the cell lysate, CTCF is co-extracted with a nucleolar protein nucleophosmin. On the other hand, the method of chromatin immunoprecipitation with antibodies to nucleophosmin was used to demonstrate the interaction of the latter with CTCF-dependent insulators *in vivo* [41]. The chromodomain-containing helicase CHD8 also binds to CTCF and CTCF-dependent insulators. In the absence of CHD8 the ICR region of the *Igf2/H19* locus ceases its insula-

tory function, although the binding of CTCF to this region is preserved [42]. The protein complexes of the CTCF-dependent insulators were demonstrated to contain a large subunit of RNA-polymerase II with the phosphorylated and dephosphorylated C-terminal domain [43].

The co-immunoprecipitation allowed determining the interaction of CTCF and the DNA-binding protein YY1. N-terminal domain of CTCF is likely to participate in the interaction. The interaction of CTCF and YY1 is required for the inactivation of one X-chromosome in order to support the active state of the second X-chromosome [44].

To date, using different methods, more proteins capable of interacting with CTCF were detected. YB1 (Y-box-binding protein 1) is a multifunctional DNA- and RNA-binding protein participating in the regulation of DNA replication and reparation, transcription, RNA processing, and capable of interacting with YY1 [45]. Kaiso is a transcription factor, capable of binding DNA regions with the increased content of methylated CpG-sites. Possibly, it is capable of replacing CTCF in case of methylation of site of the binding of the latter [46]. The transcription corepressor Sin3A [47], histone H2A.Z [22, 41], PARP [48], p68 (DDX5) [49] and other proteins [2, 50] also interact with CTCF protein.

Another important protein interacting with CTCF is cohesin. Cohesin is responsible for holding together sister chromatids required for successful mitosis and meiosis [51–53]. Cohesin is a protein complex consis-

ting of four subunits: Smc1, Smc3, Scc1 (Rad21) and Scc3 (SA1 or SA2). Four subunits form a ring structure, which, most likely, encircles and holds chromatids together [54]. Cohesin is known to be responsible for the interaction of sister chromatids, required for successful mitosis and meiosis, and to be involved in the regulation of the gene expression [55–57]. The ChIP-seq analysis demonstrated that about half of cohesin binding sites overlap with the CTCF binding sites [58–60]. Cohesin binds to the region in the C-terminal part of CTCF via its SA2 subunit [61].

The spatial proximity of the regions of interphase chromatin which simultaneously interact with cohesin and CTCF was demonstrated using 3C (chromosome conformation capture) method. The region of the chromosome between the cohesin binding sites forms a loop [56].

**The role of the transcription factor CTCF in the regulation of DNA-dependent processes.** The CTCF binding to DNA may impact the gene expression in different ways: in some cases CTCF acts as transcription activator, in others – as repressor or ensures the insulator activity.

Insulators are DNA fragments hindering the interaction of the regulatory elements between which they are located. In particular, when located between the promoter and enhancer, the insulator blocks the activating impact of the latter, whereas on the borders of the genetic construction integrated in the eukaryotic genome DNA it suppresses the position effect [62]. All known vertebrate insulators, with rare exceptions [63], bind transcription factor CTCF.

Additionally, CTCF protein participates in the inactivation of X-chromosome, the imprinting of genetic information and is likely to regulate the process of RNA splicing. The development of new methods to study the interactions between remote regions of the eukaryotic genome allowed obtaining multiple evidences of a vital role of CTCF in the formation of a three-dimensional structure of the eukaryotic genome [34, 64–68]. The understanding of the mechanisms of such numerous and various functions of CTCF protein could clear up its role in the regulation of DNA-dependent processes.

*The hypothesis on the functional domains of chromatin. Insulators.* In the late 80s of 20<sup>th</sup> century a hypothesis was put forward that all the chromatin of the eukaryotic cell is subdivided into structural/functional do-

main. It stated that a chromatin domain is a loop containing one or several genes, the ends of which are attached to the nuclear matrix. For each loop, an independent supercoiling of DNA is typical. The chromatin of one domain, regardless of the chromatin of other domains, may pass into open (transcriptionally active) or closed (inactive) conformation [69, 70].

Since the formulation of the chromatin domains hypothesis, numerous data were obtained that support, elaborate and amend this hypothesis. At present, it is established that the main factor controlling the chromatin fiber decompactization and, hence, the possibility to start transcription at the given chromosome region, is the acetylation of histones [71, 72]. The locus control regions (LCRs), *i. e.* DNA regions determining the transcriptional status of the domain, have been defined [73–75].

The study of the chicken beta-globin locus demonstrated that in the beta-globin expressing cells the chromatin of the locus is in the decompact state, the level of histone acetylation is increased. At the same time the chromatin outside the locus is in the condensed state [76–78]. The ends of the beta-globin locus interact with the nuclear matrix and are close to each other, and the locus forms a loop [79–83]. It was also demonstrated that the loops may be formed not only by bringing together functional domains ends, but also by joining specific DNA regions, for instance enhancers and promoters, inside the same domain [81, 84, 85].

Recently it has been established that the transcriptional activity of the gene depends on both the regulatory elements inside the domain and the location of this gene in a specific part of the nucleus [64–68]. The gene may also be exposed to the regulatory elements located in other domains and even on other chromosomes [86–88]. Therefore, the amended and elaborated hypothesis of the structural/functional chromatin domains remains viable.

Given the chromatin domains, there should be functional elements, which protect the genes of one domain against an impact of the regulatory elements of other domains. These functional elements of the genome are insulators [89, 90]. Insulators prevent undesirable activation or repression of genes under the impact of the environment. The undesirable gene activation by the enhancer is suppressed via blocking its effect on the promoter only in case when the insulator is located between

them. The insulator also prevents the undesirable gene repression by limiting the expansion of the condensed chromatin along the chromatin fiber [89, 91]. Some authors call insulators the border elements, since they are often located at the borders of domains (Fig. 2) [92].

The prevailing majority of insulators found in the vertebrate genomes are capable of binding the transcription factor CTCF [8, 63, 89]. CTCF plays a vital role in the formation of chromatin loops. It was demonstrated that in vertebrate beta-globin loci the CTCF binding sites located at the borders of loci are in contact with each other [80, 81, 93], and no activation of the promoters in the adjacent domains by the locus enhancers is detected in the beta-globin expressed cells. Similarly, the regulatory elements of the adjacent domains do not affect the expression of globin genes in erythroid cells [76–78, 94, 95].

The analysis of the human genome showed that the distribution of CTCF binding sites strongly correlates with the density of the genes, but weakly correlates with the lengths of chromosomes [24]. Despite the dependence of the number of CTCF binding sites on the number of genes, the great part of these sites (46 %) is located very far from promoters, 48 kbp on the average [24, 96]. It distinguishes the distribution of CTCF binding sites from the distribution of sites of the majority of other transcription factors and is in good agreement with the insulator function of CTCF.

The genome regions depleted of the CTCF binding sites mostly contain the gene families with the shared regulation of transcription, whereas the domains enriched for the CTCF binding sites contain the genes with alternatively regulated promoters. These observations are also in agreement with the insulator function of CTCF [24].

*Enhancer blocking.* The first data on the enhancer-blocking properties of CTCF binding sites were obtained using the constructions containing a promoter-driven reporter gene and an enhancer. The CTCF binding sequence was introduced between the enhancer and the promoter [8; 97]. The analysis of expression of the reporter gene in the presence and in the absence of the CTCF binding site allowed estimating the enhancer-blocking activity of this DNA fragment. The majority of CTCF binding sequences demonstrate their enhancer-blocking activity in such constructions during both

the transient transfection and the integration into the genome [98–100].

The method of estimation of the reporter gene expression at the transient transfection has some drawbacks. Normally the CTCF binding sites function as a part of chromatin, and the plasmid vector does not completely imitate the chromatin environment.

Additionally, in many cases the plasmid vector is a heterologous system where the enhancer, promoter, CTCF binding site and the reporter gene may originate from different organisms.

The mechanism of function of the CTCF-dependent insulator was studied using the artificial minichromosome containing the enhancer from human beta-globin locus LCR and the epsilon-globin gene with its own promoter [101]. The CTCF-dependent chicken 5'HS4 insulator was cloned between the enhancer and the promoter of epsilon-globin gene, and for the control – close to the enhancer outside the promoter-enhancer unit. When the insulator was placed between the enhancer and promoter, the following effects were demonstrated: the CTCF-dependent enhancer-blocking effect as well as a CTCF-dependent decrease in the amount of the promoter-bound RNA-polymerase II and an increase in the amount of polymerase interacting with the enhancer and CTCF-binding sequence. This may be explained by the fact that the CTCF-containing DNA-protein complex hinders the movement of RNA-polymerase from the enhancer towards the promoter. It was also noted the CTCF-dependent reduction in the acetylation of histones H3 and H4 between the 5'HS4 insulator and the gene, including the promoter region, when the 5'HS4 insulator was located between the enhancer and the promoter [101].

The introduction of several core sequences of the chicken 5'HS4 insulator into the plasmid construction between the enhancer and the promoter resulted in a more pronounced enhancer-blocking effect compared to the introduction of one sequence [102]. The same effect is observed while placing between the enhancer and the promoter several CTCF-dependent insulators HS1 or HS2 from ICR of the murine *Igf2/H19* locus [37, 38]. Therefore, several CTCF-containing DNA-protein complexes hinder the movement of RNA-polymerase II from the enhancer to the promoter more efficiently compared to a single complex.

However, the signal transmission along the chromatin fiber hardly may be considered as the only mechanism of the enhancer functioning. Many enhancers are separated from their target promoters by millions of base pairs [103], and the enhancer and the promoter are often separated by several chromosome loci with different chromatin structures. Moreover, the activation of the promoter by the enhancer located on the other chromosome is described [86, 88].

Some authors believe that this mechanism is realized only in enhancer-like elements of prokaryotes and not in case of real eukaryotic enhancers [104].

Recently, using the 3C (chromosome conformation capture) technology and its extensions [105], the direct interaction of the enhancer with the corresponding promoter was demonstrated [106, 107]. However, it is still unknown the mechanism of promoter activation via the chromatin loop formation and bringing the enhancer and the promoter closer to each other [108]. Possibly, the promoter is drawn into the nuclear compartment with the conditions for the active transcription [108]. The ability to stabilize the chromatin loops was demonstrated for different proteins, including CTCF, by the ChIA-PET method [34, 87].

According to the decoy model [62, 109], the CTCF-binding insulators may compete with the promoters for the interaction with enhancers, *i. e.* they may trap the enhancer not allowing its binding to the promoter. It was demonstrated [34] that a significant number of CTCF binding sequences participating in remote interactions are bound to the sequences with the enhancer properties. On the other hand, if an insulator functions as a decoy, there should be some signal transmission between the enhancer and the promoter along the chromatin fiber, otherwise the decoy would have the same effect regardless of its position relative to the enhancer which is in contradiction to the insulator definition. The 3C method was used in beta-globin loci of vertebrates to show that the enhancers and promoters are in close physical proximity [80], which is the evidence of the enhancer functioning via the combined tracking-looping mechanism.

*Insulator as a border element.* During the introduction of the transgene-containing constructs into the genome, the transgene expression initially occurs at approximately the same level in all the cells, but later the

majority of the constructs inserted into the genome become inactive due to the chromatin condensation. This inhibition is reversible in *Drosophila* and yeast, but in vertebrates it becomes irreversible due to DNA methylation [110]. The 5'HS4 insulator, located at the 5'-end of the beta-globin locus, hinders the spreading of the adjacent area of the condensed chromatin over the whole locus.

It was demonstrated by the ChIP method [111] that some CTCF binding sites are located in the regions of the transition of the compacted chromatin to the decompacted state. These regions were detected by the change in the level of trimethylation of the histone H3 lysine 27 (H3K27me3) and acetylation of the lysine 5 of histone H2A (H2AK5ac). The modification H3K27me3 is characteristic for the compact chromatin and H2AK5ac – for the decompacted one. The portion of the CTCF binding sequences located in the regions of transition of the compacted chromatin into the decompacted state is small, but non-random distribution of these sequences was demonstrated with high significance. It evidences for the probable role of the CTCF binding sites in the separation of the chromosome domains *in vivo*.

A considerable enrichment of the border regions of so called topological domains for the CTCF binding sites was demonstrated [112]. The authors assumed that the border regions of the topological domains may represent the insulators blocking the expansion of heterochromatin. Indeed, the border elements of the topological domains in the differentiated cells correspond to the regions with the reduced level of H3K9me, characteristic for the transformation of heterochromatin into a less compact state. During the formation of the final structure of the genome domains the stabilization of the chromatin state is likely to occur after the formation of the loop structure.

The CTCF binding sequences may not be directly involved in the blocking of the heterochromatin expansion. However, the distinct division of the regions of compacted and decompacted chromatin in the genome may be a result of the formation of the loop structures by the CTCF-containing border elements.

*CTCF as a transcription regulator.* CTCF is capable of both activating and inhibiting transcription depending on the target gene. The mechanisms of the activation may be different. Firstly, CTCF may act as a trans-

cription factor promoting or hindering the formation of the initiator complex. In other cases the binding of CTCF may result in the formation/destruction of the chromatin domain with the corresponding change in the chromatin structure, which, in turn, leads to the change in the genes expression in this domain. It should be noted that it is not always easy to discriminate between these mechanisms. According to the first mechanism, gene *MYC* is likely to be repressed [5], whereas to the second – gene *PUMA* [113]. Some examples of the regulation involving CTCF are presented below.

The binding of CTCF to 5'-regulatory region of APP (amyloid  $\beta$  protein precursor) gene resulted in the activation of this gene promoter at transcription *in vitro* in the nuclear extract of HeLa cells [31, 114]. The addition of oligonucleotides, competing for CTCF binding, to the nuclear extract led to a decrease of the promoter activity. The same effect was observed when CTCF was removed from the nuclear extract via immunoprecipitation, but the addition of the CTCF protein to the depleted nuclear extract restored the activity of the promoter. The region of CTCF protein between the amino acid residues 1 and 248 is responsible for the activation [114].

It was demonstrated that the HS5-1 DNA region, hypersensitive to DNase I and located in the cluster of protocadherin genes, has properties of the enhancer [115] and, according to the chromatin immunoprecipitation data, interacts with CTCF in the murine brain cells. The deletion of this CTCF binding sequence from the genome of transgene mice leads to the reduction in the promoter activity [116]. Chromatin immunoprecipitation with subsequent mass sequencing demonstrated that CTCF and cohesin interact with the promoter of alternative forms of protocadherin alpha. The binding directly correlates with the expression of alternative isoforms [117]. The 3C method was used to demonstrate the spatial proximity of the promoters of alpha-protocadherins 4, 8 and 12, and the potential enhancers HS5-1 and HS7 [84].

The chicken lysozyme silencer consists of two modules F1 and F2. Module F2 binds to the receptor of thyroid hormones, whereas F1 contains the CTCF binding site, and both modules can suppress transcription independently of each other. When each of them binds its protein factor, their activity is synergized [6]. Pos-

sibly, the mechanism of suppression-activation of the chicken lysozyme gene involves the co-suppressor Sin3A and the histone acetylase and deacetylase complexes [47, 91, 118].

The interaction of CTCF with two CpG-islands in the first intron of *Bcl6* gene does not allow this gene to be actively expressed. These CpG-islands are methylated in some lymphomas, which prevents CTCF binding to DNA and leads to an increase in the cellular *Bcl6* mRNA level [119].

Binding CTCF to the site in the first exon of human telomerase gene *hTERT* inhibits its transcription. The methylation of CTCF binding sequence is likely to play a vital role in the regulation of this gene expression. The first exon of the telomerase gene within the plasmid construction has a suppressing effect on some promoters in the cells both expressing and not expressing telomerase. In the telomerase-expressing cells, it is observed the methylation of the CTCF binding sequence in the first exon and the absence of CTCF binding [120].

After treatment with 5-aza-2'-deoxycytidine, leading to the DNA demethylation, the binding of CTCF to the site in the first exon of the *hTERT* gene and the inhibition of its mRNA synthesis are observed in the cell lines expressing *hTERT*. The inhibition of CTCF expression using short hairpin RNAs led to an increase in the level of the *hTERT* mRNA in the cells [120; 121].

CTCF participation was also noticed in the transcription regulation of Bax [122], tumor repressor p16 [123], and the ribosomal RNA genes [124].

*The participation of CTCF protein in the imprinting of genetic information.* The gene imprinting is defined as the way of inheritance, when only one parental allele of the gene is expressed in the progeny. The choice of the allele to be expressed is determined by its origin from either the paternal or maternal organism. CTCF interacts with the DNA region that controls the imprinting (imprinting control region, ICR). The most studied example of CTCF participation in the imprinting is the regulation of the genes of the mouse *Igf2/H19* locus [37, 125]. *Igf2* (insulin-like growth factor 2) gene encodes the embryonic mitogen [37], from *H19* gene the non-coding RNA is transcribed, slowing down the fetal growth [62, 126, 127]. The expression of *Igf2* occurs only from the paternal chromosome, *H19* – only from the maternal due to the ICR located between these

genes. ICR on the maternal chromosome of mice contains two sites capable of binding CTCF and having the insulator properties. ICR on the paternal chromosome is methylated regardless of the tissue and the stage of development and is not capable of binding CTCF. Therefore, the CTCF-dependent insulators on the paternal chromosome are inactive, the promoter freely interacts with the enhancers and *Igf2* is expressed [128].

While interacting with ICR of the maternal chromosome, CTCF is likely to protect the adjacent regions of the locus, including the promoter and the inner regions of gene *H19*, from the methylation. If the maternal ICR is mutated and does not bind CTCF, the methylation spreads to the promoter and the intragenic regions of *H19*, and the expression of this gene decreases [129, 130]. Therefore, by the inhibiting of the CTCF binding, the methylation of the paternal ICR leads to the methylation of the promoter of *H19* gene and the inhibition of the expression of the paternal allele [129–131].

*CTCF and the inactivation of X-chromosome.* CTCF participates in the inactivation of the mammalian X-chromosome [132]. The X-chromosome contains the X-inactivation center (Xic) that controls the inactivation of one X-chromosome in each cell during the embryogenesis and the maintenance of its silenced state. Xic includes genes *Xist* (X-inactive specific transcript), *Tsix* (the name comes from the reverse of *Xist*, which reflects the reverse orientation of the *Tsix* gene regarding to *Xist*) and *Xite* (X-inactivation intergenic transcription element), from which the non-coding RNAs are transcribed. The X-chromosome, where the actively expressing allele of *Xist* is located, is inactivated. The transcription of *Tsix* and *Xite* genes occurs from the active X-chromosome. The transcript of *Tsix* inactivates the expression of *Xist* located on the same chromosome, and the expression of *Xite* activates the expression of *Tsix*, which leads to the inhibition of the expression of *Xist*, resulting in maintenance of the activity of the X-chromosome with transcriptionally active *Tsix* and *Xite* alleles [50, 133].

The process of X-inactivation is realized in several stages [133–136]. First, the «counting» of the number of X-chromosomes in the cell, or rather the determination of the ratio between the number of X-chromosomes and autosomes [133, 134]. The next stage is the

homologous pairing of two X-chromosomes and choice of the one to be inactivated. The interaction between X-chromosomes takes place in the Xic region, and for the pairing of X-chromosomes the fragment of Xic containing *Xite* and *Tsix* genes is required [137, 138]. This region is rich in the binding sites of CTCF as well as protein YY1 [44, 132, 139]. It was demonstrated that in the absence of CTCF in the cell the pairing of X-chromosomes is suppressed. Protein YY1, capable of interacting with CTCF, is not required at this stage [139]. A model is proposed according to which the pairing of X-chromosomes leads to irreversible transfer of protein factors like CTCF and Oct-4 from one X-chromosome (later inactive) to the other (later active) [44, 138–142]. If this model is accurate, CTCF protein is also involved in the selection of the X-chromosome to be inactivated. The last stage is just the process of X-chromosome inactivation. As a result of active expression of *Xist* gene, the *Xist* transcript covers the chromosome where it was synthesized, and the chromatin of this X-chromosome becomes compacted [134–136].

The process of X-chromosome inactivation involves the interaction of the repressor complex PRC2 with 5'-region of *Xist* gene as well as the enrichment of chromatin of this region in trimethylated lysine 27 histone H3 [143, 144]. The inactivation of X-chromosome requires the transcription of a small region of *Xist* into RepA RNA [145, 146]. RepA is likely to interact with the PRC2 complex [144]. It was demonstrated that X-chromosome containing in the *Xist* promoter the CTCF binding site with the enhanced affinity to this protein is inactivated more often than X-chromosome where this DNA fragment has lower affinity to CTCF [147]. Along with YY1, CTCF is likely to participate in the activation of the *Tsix* gene expression [44]. CTCF also ensures the functioning of insulator elements, separating the active genes on the inactivated X-chromosome from the transcriptionally non-active environment [148].

*The co-transcriptional regulation of the alternative splicing.* Recently some evidences have been found of the fact that CTCF may be involved in the co-transcriptional regulation of the alternative splicing. It was demonstrated that the interaction of CTCF with the DNA fragment located in the fifth exon of *CD45* gene leads to a more frequent inclusion of this exon into the mature mRNA. CTCF depletion in the cell by RNA-in-



terference, or methylation of the CTCF binding site in the fifth exon of *CD45* gene resulted in the inhibition of the formation of a splice form containing the fifth exon. It was shown that CTCF binding to this site leads to RNA-polymerase II pausing within the region of CTCF binding. The comparison of the genome-wide chromatin immunoprecipitation data obtained with antibodies to CTCF and RNA-polymerase II and the data of the transcriptome analysis confirmed that the CTCF protein bound to DNA detains the transcribing RNA-polymerase II [149, 150].

These data were used to suggest the hypothesis on the role of the CTCF protein in the regulation of the alternative splicing. The initial stages of RNA splicing take place during the transcription, and the preferential development of a splice form may depend on the elongation rate. If in the transcript, after the occurrence of a relatively weak splice-site, but prior to the occurrence of a more strong one, the RNA-polymerase II is detained by the CTCF protein, bound to DNA, a weaker splice-site will gain an advantage. Thus, the CTCF protein may affect the elongation rate and regulate the alternative splicing [149, 151].

**Conclusions.** The main functions of CTCF, studied thus far, are as follows:

- the direct regulation of transcription of some genes;
- the organization of the domain structure of chromatin;
- the organization of insulators (enhancer-blocking and border elements);
- the participation in the genome imprinting and inactivation of X-chromosome;
- the regulation of the alternative splicing.

In most cases the accurate mechanism of these functions is yet to be studied.

A great number and variety of the CTCF binding sites, the multitude of its functions and the fact that CTCF is present only in relatively developed metazoa and is absent in plants and protozoa, suggest that its main task is to regulate the organism development and to organize the cellular genome so that it could be a constituent of a multicellular organism.

**Funding.** The work of the authors was supported by the grant program for the leading scientific schools of Russia (Project NSh\_1674.2012.4), the RAS Molecular and Cellular Biology program and the Russian

Foundation for Basic Research (Projects 10-04-01365, 10-04-01472).

Фактор транскрипції CTCF і організація геному ссавців

О. С. Котова, С. Б. Акопов, Є. Д. Свердлов, Л. Г. Ніколаєв

Резюме

*Фактор транскрипції CTCF вважають одним з основних учасників різних мереж регуляції генів, з-поміж яких активація і репресія транскрипції, утворення незалежно функціонуючих доменів хроматину, регуляція імпринтингу тощо. Секвенування геномів людини та інших організмів дозволяє виявляти геномний розподіл сайтів зв'язування CTCF та ідентифікувати CTCF-залежні регуляторні елементи, до яких належать інсулятори. В огляді підсумовано нові дані з функціонування CTCF у рамках гіпотези участі петельних доменів хроматину у великомасштабній регуляції активності геному. Фундаментальні властивості CTCF дозволяють йому діяти як регулятор транскрипції, інсуляторний білок, а також як розподілений по геному прикордонний елемент, здатний залучати різні фактори, які з'являються у відповідь на різноманітні зовнішні та внутрішні чинники і таким чином виконувати свої сигнал-специфічні функції.*

*Ключові слова:* фактор транскрипції CTCF, хроматин, регуляція транскрипції.

Фактор транскрипции CTCF и организация генома млекопитающих

Е. С. Котова, С. Б. Акопов, Е. Д. Свердлов, Л. Г. Николаев

Резюме

*Фактор транскрипции CTCF считается одним из основных участников различных сетей регуляции генов, включая активацию и репрессию транскрипции, образование независимо функционирующих доменов хроматина, регуляцию импринтинга и т. д. Секвенирование геномов человека и других организмов позволяет выявить геномное распределение сайтов связывания CTCF и идентифицировать CTCF-зависимые регуляторные элементы, включая инсуляторы. В обзоре суммированы новые данные по функционированию CTCF в рамках гипотезы участия петельных доменов хроматина в крупномасштабной регуляции активности генома. Фундаментальные свойства CTCF позволяют ему действовать как регулятор транскрипции, инсуляторный белок, а также как распределенный по геному пограничный элемент, способный привлечь различные факторы, появляющиеся в ответ на разнообразные внешние и внутренние воздействия и таким образом осуществлять свои сигнал-специфичные функции.*

*Ключевые слова:* фактор транскрипции CTCF, хроматин, регуляция транскрипции.

## REFERENCES

1. Holwerda SJ, de Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci.* 2013;**368**(1620):20120369.
2. Nikolaev LG, Akopov SB, Didych DA, Sverdlov ED. Vertebrate protein CTCF and its multiple roles in a large-scale regulation of genome activity. *Curr Genomics.* 2009;**10**(5):294–302.
3. Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell.* 2009;**137**(7):1194–211.

4. Lobanenko VV, Gudvin GG. CCCTC-binding protein: a new nuclear protein factor which interaction with 5'-flanking sequence of chicken c-myc oncogene correlates with repression of the gene. *Dokl Akad Nauk SSSR*. 1989;**309**(3):741–5.
5. Lobanenko VV, Nicolas RH, Adler VV, et al. 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*. 1990;**5**(12):1743–53.
6. Baniahmad A, Steiner C, Kohne AC, Renkawitz R. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell*. 1990;**61**(3):505–14.
7. Burcin M, Arnold R, Lutz M, et al. Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. *Mol Cell Biol*. 1997;**17**(3):1281–8.
8. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*. 1999;**98**(3):387–96.
9. Klenova EM, Nicolas RH, Paterson HF, et al. 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**(12):7612–24.
10. Robinett CC, O'Connor A, Dunaway M. The repeat organizer, a specialized insulator element within the intergenic spacer of the *Xenopus* rRNA genes. *Mol Cell Biol*. 1997;**17**(5):2866–75.
11. Filippova GN, Fagerlie S, Klenova EM, et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol*. 1996;**16**(6):2802–13.
12. Ohlsson R, Renkawitz R, Lobanenko V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet*. 2001;**17**(9):520–7.
13. Klenova EM, Chermukhin IV, El-Kady A, et al. Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol Cell Biol*. 2001;**21**(6):2221–34.
14. Klenova E, Ohlsson R. Poly(ADP-ribosylation) and epigenetics. Is CTCF PART of the plot? *Cell Cycle*. 2005;**4**(1):96–101.
15. Kitchen NS, Schoenherr CJ. Sumoylation modulates a domain in CTCF that activates transcription and decondenses chromatin. *J Cell Biochem*. 2010;**111**(3):665–75.
16. MacPherson MJ, Beatty LG, Zhou W, Du M, Sadowski PD. The CTCF insulator protein is posttranslationally modified by SUMO. *Mol Cell Biol*. 2009;**29**(3):714–25.
17. Wang J, Wang Y, Lu L. De-SUMOylation of CCCTC binding factor (CTCF) in hypoxic stress-induced human corneal epithelial cells. *J Biol Chem*. 2012;**287**(15):12469–79.
18. Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci USA*. 2012;**109**(43):17507–12.
19. Heath H, Ribeiro de Almeida C, Sleutels F, et al. CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J*. 2008;**27**(21):2839–50.
20. Splinter E, Heath H, Kooren J, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 2006;**20**(17):2349–54.
21. Fedoriw AM, Stein P, Svoboda P, Schultz RM, Bartolomei MS. Transgenic RNAi reveals essential function for CTCF in *H19* gene imprinting. *Science*. 2004;**303**(5655):238–40.
22. Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;**129**(4):823–37.
23. Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008;**133**(6):1106–17.
24. Kim TH, Abdullaev ZK, Smith AD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*. 2007;**128**(6):1231–45.
25. Whittington T, Perkins AC, Bailey TL. High-throughput chromatin information enables accurate tissue-specific prediction of transcription factor binding sites. *Nucleic Acids Res*. 2009;**37**(1):14–25.
26. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*. 2011;**147**(6):1408–19.
27. Nakahashi H, Kwon KR, Resch W, et al. A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep*. 2013;**3**(5):1678–89.
28. Ziebarth JD, Bhattacharya A, Cui Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res*. 2013;**41**(Database issue):D188–94.
29. Engel N, West AG, Felsenfeld G, Bartolomei MS. Antagonism between DNA hypermethylation and enhancer-blocking activity at the *H19* DMD is uncovered by CpG mutations. *Nat Genet*. 2004;**36**(8):883–8.
30. Renda M, Baglivo I, Burgess-Beusse B, et al. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J Biol Chem*. 2007;**282**(46):33336–45.
31. Vostrov AA, Quitschke WW. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J Biol Chem*. 1997;**272**(52):33353–9.
32. Klug A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys*. 2010;**43**(1):1–21.
33. MacPherson MJ, Sadowski PD. The CTCF insulator protein forms an unusual DNA structure. *BMC Mol Biol*. 2010;**11**:101.
34. Handoko L, Xu H, Li G, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet*. 2011;**43**(7):630–8.
35. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*. 2000;**405**(6785):482–5.
36. Filippova GN, Thienes CP, Penn BH, et al. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the *DM1* locus. *Nat Genet*. 2001;**28**(4):335–43.
37. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorske JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature*. 2000;**405**(6785):486–9.
38. Kanduri C, Pant V, Loukinov D, et al. Functional association of CTCF with the insulator upstream of the *H19* gene is parent of origin-specific and methylation-sensitive. *Curr Biol*. 2000;**10**(14):853–6.
39. Pant V, Mariano P, Kanduri C, et al. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the *H19* imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev*. 2003;**17**(5):586–90.
40. Wang H, Maurano MT, Qu H, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 2012;**22**(9):1680–8.

41. 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**(2):291–8.
42. Ishihara K, Oshimura M, Nakao M. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell*. 2006;**23**(5):733–42.
43. Chernukhin I, Shamsuddin S, Kang SY, et al. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol Cell Biol*. 2007;**27**(5):1631–48.
44. Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT. Identification of a Ctfc cofactor, Yy1, for the X chromosome binary switch. *Mol Cell*. 2007;**25**(1):43–56.
45. Chernukhin IV, Shamsuddin S, Robinson AF, et al. Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *J Biol Chem*. 2000;**275**(38):29915–21.
46. De La Rosa-Velazquez IA, Rincon-Arango H, Benitez-Bribiesca L, Recillas-Targa F. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res*. 2007;**67**(6):2577–85.
47. Lutz M, Burke LJ, Barreto G, et al. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res*. 2000;**28**(8):1707–13.
48. Guastafierro T, Cecchinelli B, Zampieri M, et al. CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J Biol Chem*. 2008;**283**(32):21873–80.
49. Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev*. 2010;**24**(22):2543–55.
50. Zlatanova J, Caiqfa P. CCCTC-binding factor: to loop or to bridge. *Cell Mol Life Sci*. 2009;**66**(10):1647–60.
51. Hirano T. At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol*. 2006;**7**(5):311–22.
52. Ishiguro K, Watanabe Y. Chromosome cohesion in mitosis and meiosis. *J Cell Sci*. 2007;**120**(Pt 3):367–9.
53. Peters JM, Tedeschi A, Schmitz J. The cohesin complex and its roles in chromosome biology. *Genes Dev*. 2008;**22**(22):3089–114.
54. Haering CH, Lowe J, Hochwagen A, Nasmyth K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell*. 2002;**9**(4):773–88.
55. Degner SC, Verma-Gaur J, Wong TP, et al. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci U S A*. 2011;**108**(23):9566–71.
56. Hadjur S, Williams LM, Ryan NK, et al. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*. 2009;**460**(7253):410–3.
57. Kagey MH, Newman JJ, Bilodeau S, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010;**467**(7314):430–5.
58. Parelho V, Hadjur S, Spivakov M, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*. 2008;**132**(3):422–33.
59. Rubio ED, Reiss DJ, Welch PL, et al. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci USA*. 2008;**105**(24):8309–14.
60. Wendt KS, Yoshida K, Itoh T, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*. 2008;**451**(7180):796–801.
61. Xiao T, Wallace J, Felsenfeld G. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol*. 2011;**31**(11):2174–83.
62. Herold M, Bartkuhn M, Renkawitz R. CTCF: insights into insulator function during development. *Development*. 2012;**139**(6):1045–57.
63. Magdinier F, Yusufzai TM, Felsenfeld G. Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor alpha and Dad1 genes. *J Biol Chem*. 2004;**279**(24):25381–9.
64. Chambeyron S, Bickmore WA. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*. 2004;**18**(10):1119–30.
65. Ferrai C, Xie SQ, Luraghi P, et al. Poised transcription factories prime silent uPA gene prior to activation. *PLoS Biol*. 2010;**8**(1):e1000270.
66. Mahy NL, Perry PE, Bickmore WA. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol*. 2002;**159**(5):753–63.
67. Osborne CS, Chakalova L, Brown KE, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*. 2004;**36**(10):1065–71.
68. Wei Z, Huang D, Gao F, et al. Biological implications and regulatory mechanisms of long-range chromosomal interactions. *J Biol Chem*. 2013;**288**(31):22369–77.
69. Bodnar JW. A domain model for eukaryotic DNA organization: a molecular basis for cell differentiation and chromosome evolution. *J Theor Biol*. 1988;**132**(4):479–507.
70. Goldman MA. The chromatin domain as a unit of gene regulation. *Bioessays*. 1988;**9**(2–3):50–5.
71. Anguita E, Johnson CA, Wood WG, Turner BM, Higgs DR. Identification of a conserved erythroid specific domain of histone acetylation across the alpha-globin gene cluster. *Proc Natl Acad Sci U S A*. 2001;**98**(21):12114–9.
72. Krajewski WA, Becker PB. Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. *Proc Natl Acad Sci USA*. 1998;**95**(4):1540–5.
73. Forrester WC, Epner E, Driscoll MC, et al. A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes Dev*. 1990;**4**(10):1637–49.
74. Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*. 1987;**51**(6):975–85.
75. Razin SV, Farrell CM, Recillas-Targa F. Genomic domains and regulatory elements operating at the domain level. *Int Rev Cytol*. 2003;**226**:63–125.
76. Felsenfeld G. Chromatin structure and the expression of globin-encoding genes. *Gene*. 1993;**135**(1–2):119–24.
77. Hebbes TR, Clayton AL, Thorne AW, Crane-Robinson C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J*. 1994;**13**(8):1823–30.
78. Litt MD, Simpson M, Recillas-Targa F, Prioleau MN, Felsenfeld G. Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *EMBO J*. 2001;**20**(9):2224–35.
79. Kurukuti S, Tiwari VK, Tavoosidana G, et al. 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**(28):10684–9.
80. 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**(2):190–4.

81. *Ulianov SV, Gavrilov AA, Razin SV.* Spatial organization of the chicken beta-globin gene domain in erythroid cells of embryonic and adult lineages. *Epigenetics Chromatin.* 2012;**5**(1):16.
82. *Wang L, Di LJ, Lv X, et al.* Inter-MAR association contributes to transcriptionally active looping events in human beta-globin gene cluster. *PLoS One.* 2009;**4**(2):e4629.
83. *Yusufzai TM, Felsenfeld G.* The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci USA.* 2004;**101**(23):8620–4.
84. *Guo Y, Monahan K, Wu H, et al.* CTCF/cohesin-mediated DNA looping is required for protocadherin  $\alpha$  promoter choice. *Proc Natl Acad Sci U S A.* 2012;**109**(51):21081–6.
85. *Mitchell JA, Fraser P.* Transcription factories are nuclear sub-compartments that remain in the absence of transcription. *Genes Dev.* 2008;**22**(1):20–5.
86. *Hu Q, Kwon YS, Nunez E, et al.* Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc Natl Acad Sci U S A.* 2008;**105**(49):19199–204.
87. *Li G, Ruan X, Auerbach RK, et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell.* 2012;**148**(1–2):84–98.
88. *Patel B, Kang Y, Cui K, et al.* Aberrant TAL1 activation is mediated by an interchromosomal interaction in human T-cell acute lymphoblastic leukemia. *Leukemia.* 2014;**28**(2):349–61.
89. *West AG, Gaszner M, Felsenfeld G.* Insulators: many functions, many mechanisms. *Genes Dev.* 2002;**16**(3):271–88.
90. *Wallace JA, Felsenfeld G.* We gather together: insulators and genome organization. *Curr Opin Genet Dev.* 2007;**17**(5):400–7.
91. *Dunn KL, Zhao H, Davie JR.* The insulator binding protein CTCF associates with the nuclear matrix. *Exp Cell Res.* 2003;**288**(1):218–23.
92. *Gohl D, Aoki T, Blanton J, Shanower G, Kappes G, Schedl P.* Mechanism of chromosomal boundary action: roadblock, sink, or loop? *Genetics.* 2011;**187**(3):731–48.
93. *Chien R, Zeng W, Kawachi S, et al.* Cohesin mediates chromatin interactions that regulate mammalian  $\beta$ -globin expression. *J Biol Chem.* 2011;**286**(20):17870–8.
94. *Bulger M, Bender MA, van Doorninck JH, et al.* Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters. *Proc Natl Acad Sci U S A.* 2000;**97**(26):14560–5.
95. *Schubeler D, Francastel C, Cimborra DM, Reik A, Martin DI, Groudine M.* Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev.* 2000;**14**(8):940–50.
96. *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 USA.* 2007;**104**(17):7145–50.
97. *Recillas-Targa F, Bell AC, Felsenfeld G.* Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *Proc Natl Acad Sci USA.* 1999;**96**(25):14354–9.
98. *Akopov SB, Ruda VM, Batrak VV, et al.* Identification, genome mapping, and CTCF binding of potential insulators within the FXVD5-COX7A1 locus of human chromosome 19q13.12. *Mamm Genome.* 2006;**17**(10):1042–9.
99. *Didych DA, Kotova ES, Akopov SB, Nikolaev LG, Sverdlov ED.* DNA fragments binding CTCF *in vitro* and *in vivo* are capable of blocking enhancer activity. *BMC Res Notes.* 2012;**5**:178.
100. *Vetchinova AS, Akopov SB, Chernov IP, Nikolaev LG, Sverdlov ED.* Two-dimensional electrophoretic mobility shift assay: identification and mapping of transcription factor CTCF target sequences within an FXVD5-COX7A1 region of human chromosome 19. *Anal Biochem.* 2006;**354**(1):85–93.
101. *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**(16):4903–19.
102. *Chung JH, Bell AC, Felsenfeld G.* Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci USA.* 1997;**94**(2):575–80.
103. *Harmston N, Lenhard B.* Chromatin and epigenetic features of long-range gene regulation. *Nucleic Acids Res.* 2013;**41**(15):7185–99.
104. *Kulaeva OI, Nizovtseva EV, Polikanov YS, Ulianov SV, Studitsky VM.* Distant activation of transcription: mechanisms of enhancer action. *Mol Cell Biol.* 2012;**32**(24):4892–7.
105. *Simonis M, Kooren J, de Laat W.* An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods.* 2007;**4**(11):895–901.
106. *Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W.* Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell.* 2002;**10**(6):1453–65.
107. *Tsytsykova AV, Rajsbaum R, Falvo JV, Ligeiro F, Neely SR, Goldfeld AE.* Activation-dependent intrachromosomal interactions formed by the TNF gene promoter and two distal enhancers. *Proc Natl Acad Sci USA.* 2007;**104**(43):16850–5.
108. *Krivega I, Dean A.* Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev.* 2012;**22**(2):79–85.
109. *Geyer PK.* The role of insulator elements in defining domains of gene expression. *Curr Opin Genet Dev.* 1997;**7**(2):242–8.
110. *Pikaart MJ, Recillas-Targa F, Felsenfeld G.* Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev.* 1998;**12**(18):2852–62.
111. *Cuddapah S, Jothi R, Schonets DE, Roh TY, Cui K, Zhao K.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* 2009;**19**(1):24–32.
112. *Dixon JR, Selvaraj S, Yue F, et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012;**485**(7398):376–80.
113. *Gomes NP, Espinosa JM.* Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding. *Genes Dev.* 2010;**24**(10):1022–34.
114. *Vostrov AA, Taheny MJ, Quitschke WW.* A region to the N-terminal side of the CTCF zinc finger domain is essential for activating transcription from the amyloid precursor protein promoter. *J Biol Chem.* 2002;**277**(2):1619–27.
115. *Ribich S, Tasic B, Maniatis T.* Identification of long-range regulatory elements in the protocadherin- $\alpha$  gene cluster. *Proc Natl Acad Sci USA.* 2006;**103**(52):19719–24.
116. *Kehayova P, Monahan K, Chen W, Maniatis T.* Regulatory elements required for the activation and repression of the protocadherin- $\alpha$  gene cluster. *Proc Natl Acad Sci USA.* 2011;**108**(41):17195–200.
117. *Monahan K, Rudnick ND, Kehayova PD, et al.* Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin- $\alpha$  gene expression. *Proc Natl Acad Sci USA.* 2012;**109**(23):9125–30.
118. *Lutz M, Burke LJ, LeFevre P, et al.* Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor. *EMBO J.* 2003;**22**(7):1579–87.
119. *Lai AY, Fatemi M, Dhasarathy A, et al.* DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *J Exp Med.* 2010;**207**(9):1939–50.

120. Renaud S, Loukinov D, Bosman FT, Lobanenkova V, Benhattar J. CTCF binds the proximal exonic region of hTERT and inhibits its transcription. *Nucleic Acids Res.* 2005;**33**(21):6850–60.
121. Renaud S, Loukinov D, Abdullaev Z, et al. Dual role of DNA methylation inside and outside of CTCF-binding regions in the transcriptional regulation of the telomerase hTERT gene. *Nucleic Acids Res.* 2007;**35**(4):1245–56.
122. Mendez-Catala CF, Gretton S, Vostrov A, et al. A novel mechanism for CTCF in the epigenetic regulation of Bax in breast cancer cells. *Neoplasia.* 2013;**15**(8):898–912.
123. Rayess H, Wang MB, Srivatsan ES. Cellular senescence and tumor suppressor gene p16. *Int J Cancer.* 2012;**130**(8):1715–25.
124. Huang K, Jia J, Wu C, et al. Ribosomal RNA gene transcription mediated by the master genome regulator protein CCCTC-binding factor (CTCF) is negatively regulated by the condensin complex. *J Biol Chem.* 2013;**288**(36):26067–77.
125. Kanduri C, Holmgren C, Pilartz M, et al. The 5' flank of mouse H19 in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication. *Curr Biol.* 2000;**10**(8):449–57.
126. Gabory A, Jammes H, Dandolo L. The H19 locus: role of an imprinted non-coding RNA in growth and development. *Bioessays.* 2010;**32**(6):473–80.
127. Leighton PA, Saam JR, Ingram RS, Stewart CL, Tilghman SM. An enhancer deletion affects both H19 and Igf2 expression. *Genes Dev.* 1995;**9**(17):2079–89.
128. Lewis A, Murrell A. Genomic imprinting: CTCF protects the boundaries. *Curr Biol.* 2004;**14**(7):R284–6.
129. Schoenherr CJ, Levorse JM, Tilghman SM. CTCF maintains differential methylation at the Igf2/H19 locus. *Nat Genet.* 2003;**33**(1):66–9.
130. Szabo PE, Tang SH, Silva FJ, Tsark WM, Mann JR. Role of CTCF binding sites in the Igf2/H19 imprinting control region. *Mol Cell Biol.* 2004;**24**(11):4791–800.
131. Gao ZH, Suppola S, Liu J, Heikkila P, Janne J, Voutilainen R. Association of H19 promoter methylation with the expression of H19 and IGF-II genes in adrenocortical tumors. *J Clin Endocrinol Metab.* 2002;**87**(3):1170–6.
132. Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT. CTCF, a candidate trans-acting factor for X-inactivation choice. *Science.* 2002;**295**(5553):345–7.
133. Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* 2009;**23**(16):1831–42.
134. Avner P, Heard E. X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet.* 2001;**2**(1):59–67.
135. Clerc P, Avner P. Random X-chromosome inactivation: skewing lessons for mice and men. *Curr Opin Genet Dev.* 2006;**16**(3):246–53.
136. Erwin JA, Lee JT. New twists in X-chromosome inactivation. *Curr Opin Cell Biol.* 2008;**20**(3):349–55.
137. Bacher CP, Guggiari M, Brors B, et al. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol.* 2006;**8**(3):293–9.
138. Xu N, Tsai CL, Lee JT. Transient homologous chromosome pairing marks the onset of X inactivation. *Science.* 2006;**311**(5764):1149–52.
139. Xu N, Donohoe ME, Silva SS, Lee JT. Evidence that homologous X-chromosome pairing requires transcription and Ctfc protein. *Nat Genet.* 2007;**39**(11):1390–6.
140. Nicodemi M, Panning B, Prisco A. A thermodynamic switch for chromosome colocalization. *Genetics.* 2008;**179**(1):717–21.
141. Nicodemi M, Prisco A. Self-assembly and DNA binding of the blocking factor in X chromosome inactivation. *PLoS Comput Biol.* 2007;**3**(11):e210.
142. Nicodemi M, Prisco A. Symmetry-breaking model for X-chromosome inactivation. *Phys Rev Lett.* 2007;**98**(10):108104.
143. Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell.* 2006;**21**(5):617–28.
144. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science.* 2008;**322**(5902):750–6.
145. Hoki Y, Kimura N, Kanbayashi M, et al. A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development.* 2009;**136**(1):139–46.
146. Wutz A, Rasmussen TP, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet.* 2002;**30**(2):167–74.
147. Pugacheva EM, Tiwari VK, Abdullaev Z, et al. 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.* 2005;**14**(7):953–65.
148. Filippova GN, Cheng MK, Moore JM, et al. Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev Cell.* 2005;**8**(1):31–42.
149. Shukla S, Kavak E, Gregory M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature.* 2011;**479**(7371):74–9.
150. Wada Y, Ohta Y, Xu M, et al. A wave of nascent transcription on activated human genes. *Proc Natl Acad Sci USA.* 2009;**106**(43):18357–61.
151. Kornblihtt AR. CTCF: from insulators to alternative splicing regulation. *Cell Res.* 2012;**22**(3):450–2.

Received 15.03.14