Human Molecular Genetics, 2005, Vol. 14, No. 7 **953–965**
doi:10.1093/hmg/ddi089
Advance Access published on February 24, 2005

Familial cases of point mutations in the *XIST* promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation

Elena M. Pugacheva[1], Vijay Kumar Tiwari[2], Ziedulla Abdullaev[1], Alexander A. Vostrov[3], Patrick T. Flanagan[1], Wolfgang W. Quitschke[3], Dmitri I. Loukinov[1], Rolf Ohlsson[2,†] and Victor V. Lobanenkov[1,†,*]

[1]Molecular Pathology Section, Laboratory of Immunopathology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20852, USA, [2]Department of Development and Genetics, Uppsala University, Norbyvägen 18A, S-752 36 Uppsala, Sweden and [3]Department of Psychiatry and Behavioral Science, State University of New York at Stony Brook, Stony Brook, New York 11794-81012, USA

Received December 12, 2004; Revised February 4, 2005; Accepted February 14, 2005

The choice mechanisms that determine the future inactive X chromosome in somatic cells of female mammals involve the regulated expression of the *XIST* gene. A familial C(−43)G mutation in the *XIST* promoter results in skewing of X chromosome inactivation (XCI) towards the inactive X chromosome of heterozygous females, whereas a C(−43)A mutation found primarily in the active X chromosome results in the opposite skewing pattern. Both mutations point to the existence of a factor that might be responsible for the skewed patterns. Here we identify this factor as CTCF, a conserved protein with a 11 Zn-finger (ZF) domain that can mediate multiple sequence-specificity and interactions between DNA-bound CTCF molecules. We show that mouse and human *Xist/XIST* promoters contain one homologous CTCF-binding sequence with the matching dG-contacts, which in the human *XIST* include the −43 position within the DNase I footprint of CTCF. While the C(−43)A mutation abrogates CTCF binding, the C(−43)G mutation results in a dramatic increase in CTCF-binding efficiency by altering ZF-usage mode required for recognition of the altered dG-contacts of the mutant site. Thus, the skewing effect of the two −43C mutations correlates with their effects on CTCF binding. Finally, CTCF interacts with the *XIST/Xist* promoter only in female human and mouse cells. The interpretation that this reflected a preferential interaction with the promoter of the active *Xist* allele was confirmed in mouse fetal placenta. These observations are in keeping with the possibility that the choice of X chromosome inactivation reflects stabilization of a higher order chromatin conformation impinging on the CTCF–*XIST* promoter complex.

INTRODUCTION

To mediate dosage compensation, one of the two X chromosomes of female mammals is inactivated (1,2). This process, which is called X inactivation, is normally random in the soma of eutherian mammals whereas it is parent-of-origin-dependent in the extra-embryonic tissues (3,4). Genetic dissection has revealed that X inactivation involves a complicated set of events that culminates in the activation of *XIST* on the future inactive X chromosome (5). The *XIST* transcript

*To whom correspondence should be addressed at: Molecular Pathology Section, LIP NIAID NIH, Twinbrook I, Room 1417, MSC-8152, 5640 Fisher Lane, Rockville, MD 20852, USA. Tel: +1 301 435 1690; Fax: +1 301 402 0077; Email: vlobanenkov@niaid.nih.gov
†These authors contributed equally.

subsequently coats the chromatin fiber to irreversibly induce the inactive state along the bulk of the chosen X chromosome in somatic cells (6,7).

Although the mechanisms underlying the activation of *XIST* remain poorly understood, they are likely to involve the antisense non-coding transcript *Tsix* (8), because deletion of the *Tsix* gene or regulatory elements controlling *Tsix* expression leads to skewed X inactivation (8,9). In a more recent study, Ogawa and Lee (10) functionally characterized the 'X-inactivation intergenic transcription elements' (*Xite*) region. *Xite* is candidate enhancer element possibly regulated by tandem CTCF-binding sites in a chromatin-insulator within the *Tsix*/DXPas34 region (11). The mouse *Tsix*/DXPas34 region contains more than 40 CTCF motifs, whereas the corresponding region of the human X chromosome has more than 10 similar sites (11). X Chromosome inactivation (XCI) is not affected only by the regions of *Xist*. Further characterization of the 5′ end of the *Xist* locus revealed specific zones of non-coding transcription where a series of targeted deletions and mutations also results in a pattern of XCI skewing similar to that resulting from deletions in the *Xite* or *Tsix*/DXPas34 regions (12). Thus, there are perhaps several additional candidate regions on both sides of *Xist*, which together may define a regulated balance between sense and antisense transcription across *Xist* prior to the onset of random X inactivation.

However, a similarly skewed X inactivation could be achieved more directly by modulating the activity of the *XIST* promoter itself. Although familial *XIST* mutations are rarely found in humans, it has been nevertheless possible in a few cases to link an extreme skewing for inactivation of one X chromosome to germline mutations of a particular region in the *XIST* promoter. For example, all females belonging to two unrelated families demonstrated preferential inactivation of X chromosomes carrying identical cytosine to guanine germline substitutions at the same −43 position upstream of the *XIST* transcriptional start site C(−43)G mutation (13). While suggesting that this mutation may hit a binding site for a transcriptional factor(s) important for proper regulation of the *XIST* promoter, the value of these findings remains unclear until such a factor(s) have been identified.

Intriguingly, another substitution of adenosine at the same cytosine at −43 position of the *XIST* promoter [C(−43)A mutation], on the ring X chromosome of a 3.5-year-old girl with short stature, facial dimorphism and developmental delay, was associated with escape from X inactivation (14). Because this mutation occurs at the same dC-base as the C(−43)G mutation, but is associated preferentially with the active X chromosome, we hypothesized that the opposite choices of XCI converged on the regulation of affinity for a *trans*-acting factor specifically recognizing the *XIST* promoter. Using an experimental protein-binding screen for the region of *XIST* promoter (from +31 to −1080 position) and chromatin immunoprecipitation (ChIP) analysis, this factor is identified here as CTCF.

CTCF is a highly conserved chromatin protein already implicated in a remarkably diverse number of functions including regulation of imprinted genes in soma and of XCI in mice (11,15,16). Different functions of CTCF appear to be mediated by binding to highly dissimilar ∼50 bp-long target sequences by engaging varying combinations of its 11 Zn-fingers (ZFs) in establishing specific DNA base contacts. Because CTCF-contacting DNA base motifs have no single consensus that would allow one to reliably predict novel CTCF target sites (15), we initially utilized recombinant CTCF to map and characterize its interactions with the wild-type and mutant human *XIST* promoters *in vitro*. In addition, we studied binding of CTCF to the wild-type human and mouse *XIST/Xist* promoters *in vivo*. We demonstrate here that CTCF binds with a similar, relatively modest affinity to the same region of sequence homology in mouse, rabbit, horse and human *XIST* wild-type promoters (17), although the C(−43)G and C(−43)A mutations display increased and absence of CTCF binding, respectively.

On the basis of the fact that CTCF efficiently interacts with itself to mediate site-specific heterodimerization of CTCF–DNA complexes (18) and the fact that CTCF is *in vivo* associated with the active *Xist* promoter of only the inactive X chromosome, we discuss the possibility that the inherited point mutations at position −43 of the *XIST* promoter may pre-empt XCI choice by facilitating the formation of a higher order chromatin conformation permissive for *XIST* transcription.

RESULTS

The mouse/human *Xist/XIST* promoters are specifically recognized by the multivalent 11 ZF DNA binding domain of CTCF

Attempts to predict CTCF-binding sequences, solely based on homology with the known targets, often results in finding of such sequences among numerous genomic regions previously tested experimentally negative for CTCF binding (19). Therefore, we carried out a direct systematic search for CTCF-binding sites by using recombinant CTCF protein an electrophoretic mobility shift assays (EMSA) with two sets of eight consecutive overlapping DNA probes. One set of EMSA probes spanned the *XIST* promoter of human origin (Fig. 1B), whereas a similar set of probes covered the promoter from *Mus musculus domesticus* (Fig. 1C).

CTCF sites can be missed by conventional EMSA with DNA probes made as 20–60 bp long double-stranded oligonucleotides. These probes are too short to accommodate binding of the 11 ZF domain that generates ∼50–60 bp long DNAse I footprints on each strand with each individual ZF (plus the linker) requiring up to 5–6 bases of DNA for the efficient binding (20). Therefore, each set of ^{32}P-end-labeled DNA probes used in Figure 1 was designed to include a 120–250 bp long overlapping DNA fragments. In addition, the overlapping of DNA fragments were designed such that an additional flanking DNA around ∼50 bp of a putative CTCF-binding sequence would become included in one of two consecutive fragments.

In both screens (Fig. 1B and C), sequence-specific binding of CTCF was detected only to DNA fragment no. 8, which in both human and mouse species contains the *XIST*/*Xist* transcriptional start site. The DNA sequences of the binding fragments showed ∼78% overall homology in several

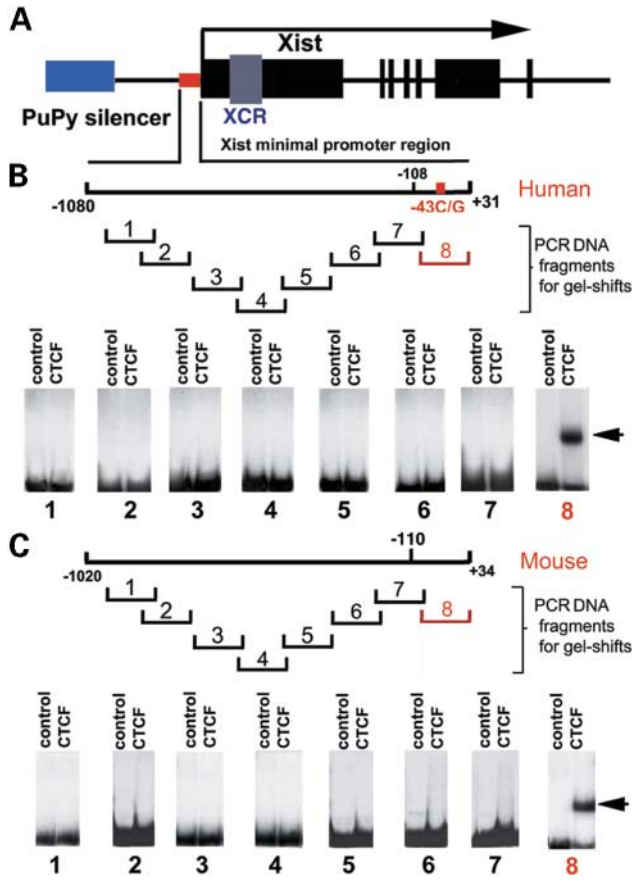

Figure 1. CTCF binding to the human *XIST* promoter region containing C(−43)G or C(−43)A mutations associated with familial skewing of the XCI, and to the similar region in the mouse *Xist* promoter. **(A)** Schematic map of the human *XIST* locus with the region of minimal promoter characterized by Hendrich *et al.* (17) and other functionally relevant regulatory elements reviewed by Heard (36). **(B)** Depiction of overlapping DNA fragments used in the EMSA for the search of CTCF-binding sites in human *XIST* minimal promoter. Below the map, eight panels from the same EMSA gel show the outcome of the binding reactions with each ^{32}P-DNA probe and *in vitro* translated (IVT) luciferase protein (control) or IVT 11 ZF of CTCF protein (CTCF). **(C)** A similar map to (B), but for DNA fragments from the mouse *Xist* promoter, prepared and used for EMSA as described earlier. Note that only sequences present in the *XIST/Xist* fragments 8 (highlighted in red) contain a single CTCF-binding site and no non-specific background as suggested by the lack of additional shifted bands. The arrowheads point to the same-mobility specific DNA–protein complexes (panels 8). All pictures taken from the same film exposed to dry EMSA gel are shown altogether by placing images of each free ^{32}P-DNA probe (F) to the bottom of each panel.

species, including human, mouse, rabbit and horse (17). A multi species sequence alignment of mammalian *XIST* promoter presented by Hendrich *et al.* (17) revealed within fragment 8 several near-identical nucleotide motifs, or phylogenetic footprints, that are the likely binding sites for conserved transcriptional factors (21). Because CTCF alone cannot account for all of these potential binding sites, it is possible that several other conserved factors in addition to CTCF may interact with the *XIST* promoter region around and immediately upstream of the transcriptional start site.

CTCF is preferentially associated *in vivo* with the active *XIST* promoter on the inactive X chromosome in human somatic cells, and with the homologous region of mouse *Xist* allele of the paternally inherited inactive X chromosome in female mouse placentas

To ascertain the *in vivo* relevance of our *in vitro* observations, we examined the interaction between CTCF and the *XIST* promoter of the inactive X in human somatic cells of female and male origin. Figure 2A shows the results of the ChIP assays with primary cultures of human mammary epithelial cells (HMEC) and normal human dermal fibroblast (NHDF) derived from healthy adult men and women. As positive control for *in vivo* CTCF binding, we used human *c-MYC* oncogene 5′-insulator site (N-site), known to be constitutively occupied by CTCF in normal cells (15,22). This experiment revealed that only female human cells, regardless of their origin, had CTCF bound to *Xist* promoter *in vivo*. In contrast, the control (N-site) has been pulled down equally well by CTCF antibody from both male and female cells. This result could be extended to freshly isolated peripheral blood leukocytes (PBLs) from healthy male and female donors for ChIP analyses. Figure 2B illustrates a 6–7-fold enrichment of DNA ChIPed with CTCF antibody from female PBLs, whereas there was no enrichment in ChIP analyses of male PBLs.

To examine more closely the possibility of a specific *in vivo* association between CTCF and the *Xist* allele of only inactive X chromosome in mammalian female cells, we turned our attention to the female mouse placenta, a tissue that exhibits selective inactivation of the paternal X chromosome. Dispersed, formaldehyde-crosslinked placental cells were subjected to ChIP analysis using affinity-purified rabbit polyclonal antibodies against the C-terminal portion of CTCF (23). Figure 2C shows results of the ChIP assays comparing a positive, internal control, the maternal *H19* ICR allele that interacts constitutively with CTCF *in vivo* (23), with the *Xist* promoter region positive for CTCF binding in the *in vitro* EMSA (Fig. 1C). These results show that the CTCF antibody pulled down both the *H19* ICR and the *Xist* promoter fragments from female placenta, whereas only the *H19* ICR could be recovered in CTCF ChIP material derived from the male placenta.

These results agree well with the ChIP data from the human cells, reinforcing the possibility that CTCF was associated with the *Xist* promoter sequence of only the inactive X chromosome. To directly examine this contention, CTCF ChIP DNA samples derived from female placenta of interspecific *M. musculus* and *M. spretus* crosses were sequenced directly. This approach exploited an A/C polymorphism at position −88 upstream of the transcriptional start site of the *Xist* gene in *M. musculus* (C57BL/6J) compared with *M. spretus* allowing us to distinguish a maternally inherited allele from a paternally inherited allele in mouse placentas from female F1 mice. Figure 2D shows that the CTCF antibody pulled down only the paternally inherited allele (*M. spretus*), confirming our supposition that CTCF is associated only with the *Xist* promoter of the inactive X chromosome in female mouse placentas and, by extrapolation, also to female humans.

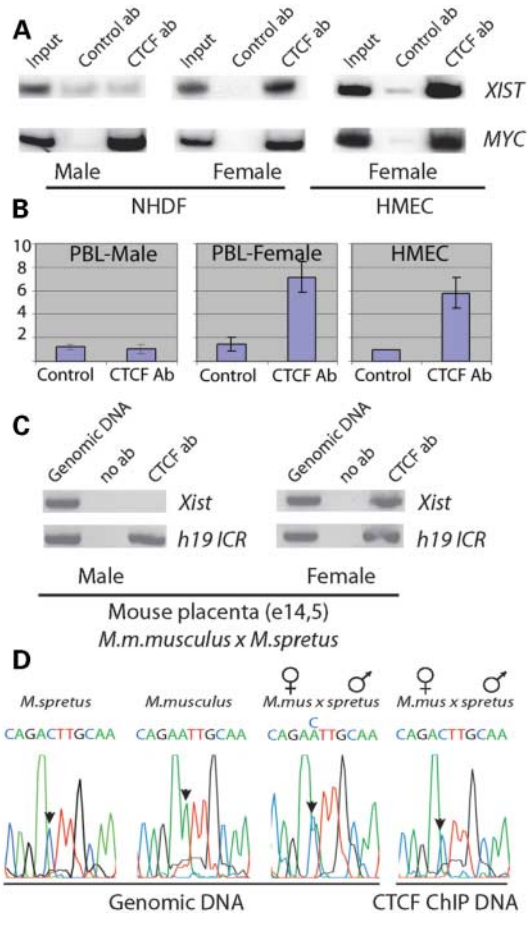

Figure 2. CTCF *in vivo* occupancy on the active wild-type *Xist*/*XIST* promoter. **(A)** ChIP of male and female NHDF and female HMEC. For each ChIP, sonicated chromatin was immunoprecipitated with either anti-CTCF antibodies or control (pre-immune) antibodies and then analyzed by PCR as described (37,38) with the pairs of primers for the human *XIST* promoter CTCF site *(XIST)* and for the Myc-N insulator site *(MYC)* (15). Each panel shows representative results of three independent PCR analyses of an IP. **(B)** Real-time PCR analysis of ChIP DNA from peripheral blood lymphocytes (PBL) of male and female donors and HMEC. This ChIP has been performed using mixture of nine anti-CTCF monoclonal antibodies (Supplementary Material). The data were similar to that obtained with polyclonal antibodies (data not shown) but monoclonals gave far superior reproducibility and thus were much better to use for quantitative PCR. Each data point indicates the average of three independent PCR analyses of an IP with the standard deviation shown by the error bars. Fold enrichment is calculated as described in Materials and Methods. **(C)** CTCF-specific ChIP analyses were performed in male and female fetal *M.m. musculus* × *M. spretus* placentas (E14.5). On the left (Male) and on the right (Female) panels, which depict results of one from at least three representative ChIP experiments, the degree of immunopurification was visualized on DNA gels used to monitor the yield from the semi-quantitative hot PCR reactions designed and performed for the control region *(H19 ICR)* and the test *(Xist)* sites essentially as described earlier for the H*19 ICR* by Kanduri *et al.* (23). A non-immune sera (no ab) was used as a control for a non-specific background generated in the ChIP procedures with CTCF antibodies (CTCF ab), whereas partially degraded total mouse DNA (Genomic DNA) was used to verify quantitative PCR conditions preliminary adjusted separately for the ChIP assays with two different primer pairs for the mouse *h19 ICR* and *Xist* CTCF-binding regions. **(D)** Direct sequencing analysis of DNA from CTCF-containing chromatin fractions, obtained by the ChIP with female placentas of the E14.5 progeny from inter-crossed hybrids between *M. musculus* and *M. spretus.* The arrows on the sequencing readout diagrams point to the C/A polymorphism, which served to distinguish between CTCF-unbound maternally inherited allele from CTCF-bound paternally inherited allele.

Germline mutations of the CTCF target sequence in the *XIST* promoter in families with skewed X inactivation

The cytosine at position −43 of the *XIST* promoter may play an important role in the X inactivation process: X chromosomes harboring a cytosine to guanine C(−43)G mutation at this position were preferentially inactivated in females of two unrelated families (13). Conversely, a cytosine to adenosine mutation at the same position, C(−43)A on the ring X chromosome of a young girl, was associated with escape from X inactivation (14). Because these mutations are located in the middle of fragment 8 of the *XIST* promoter, which is conserved among mammals and contains a target for CTCF, it seemed possible that these mutations might affect CTCF binding.

To examine this possibility, we used EMSA to test a 140 bp fragment of the human *XIST* promoter from −108 to +31 containing a wild-type sequence or the same fragment with either the C(−43)G or the C(−43)A mutation (Fig. 3A). While the C(−43)G mutation dramatically enhanced the affinity of CTCF binding when compared with the wild-type probe, CTCF binding was abolished by the C(−43)A mutation (Fig. 3B). This result observed with *in vitro* translated CTCF protein was also seen with nuclear extracts from various cells (Fig. 3B and data not shown). The specificity of CTCF binding in nuclear extracts was confirmed by a super-shift EMSA with CTCF antibodies (Fig. 3B). Promoter fragments with either the wild-type sequence or the C(−43)G mutation were efficiently bound by CTCF and the resulting CTCF–DNA complexes were super-shifted with antibodies against CTCF (Fig. 3B). In sharp contrast, no CTCF–DNA interactions could be detected with a fragment containing the C(−43)A mutation or the 'dG-contacts mutation' shown in Figure 3. The latter mutation was designed to eliminate CTCF binding by mutating a string of six major CTCF-contacting guanines (Fig. 3A). Those guanines were identified as bases involved in recognition of *XIST* by the 11 ZF domain of CTCF by DMS methylation interference assay (Fig. 5). Thus, the C(−43)A mutation eliminates CTCF binding as efficiently as the designed dG-contacts mutation that destroys the CTCF-recognition sequence.

Figure 4 presents data from competitive EMSA carried out to compare binding affinities of CTCF to the wild-type and to the C(−43)G-mutated *XIST* promoter regions. A ∼100-fold excess of unlabeled cold DNA fragment bearing the C(−43)G mutation to the labeled wild-type DNA probe completely abolished formation of the normal CTCF DNA complex (Fig. 4A, lane 3). Under similar conditions, the fragment with the C(−43)A mutation had no effect on CTCF binding to the wild-type *XIST* site (Fig. 4A, lane 8). Similar experiments performed with a labeled EMSA probe containing the C(−43)G mutation showed that an excess of the wild-type competitor DNA fragment resulted in only a marginally decreased formation of the complex produced by the probe with the mutation (Fig. 4A, lane 13). No competitive effect on this complex was observed with the cold C(−43)A DNA (Fig. 4A, lane 16). Moreover, in EMSA experiments with increasing amounts of cold DNA fragment containing a high-affinity CTCF-binding site, DMD4 from the mouse *H19* ICR (23), the labeled mutant C(−43)G *XIST* probe was

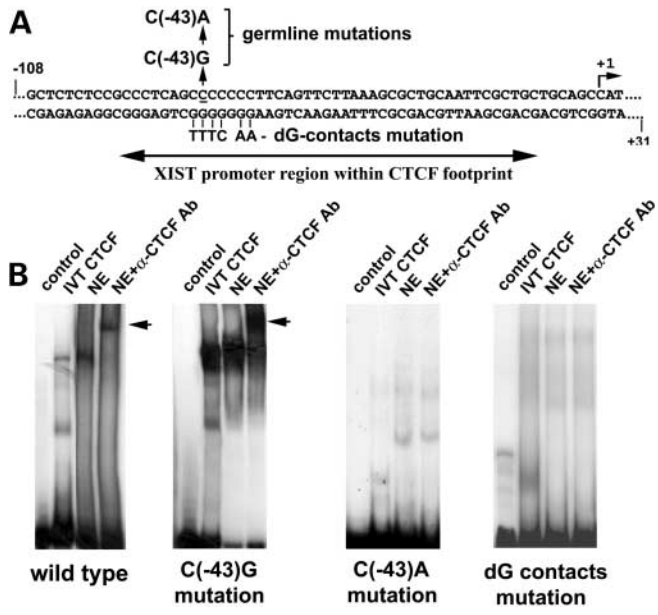

Figure 3. Effects of the C(−43)G, C(−43)A and dG-contacts mutations on CTCF binding to the *XIST* promoter. **(A)** The wild-type −108 to +31 sequence of the human *XIST* in the double-stranded DNA fragments used for EMSA. The positions and type of the natural skewing germline mutations are indicated by the arrows, whereas the dG-contacts mutation (made as explained in the text within the CTCF-footprint depicted by the double-sided arrow underneath of the sequence) is shown by bars pointing to each mutated guanine. **(B)** From left to right, the results of EMSA are shown for the wild-type, C(−43)G, C(−43)A and dG-contacts mutant *XIST* minimal promoter fragments, incubated with IVT luciferase (control), recombinant CTCF (IVT CTCF) and K562 nuclear extract without (NE) and with α-anti-CTCF antibodies (NE + α-CTCF ab). Arrows show super-shifts in the lanes with anti-CTCF antibodies, thereby confirming specific CTCF binding to both wild-type and C(−43)G mutant *XIST* promoter probes.

much more efficient in competing for CTCF than the labeled wild-type *XIST* probe. On the basis of pair-wise comparisons of the DNA-bound CTCF bands in EMSA lanes 4–7 of Figure 4A with lanes 12–15, it is seen that an estimated 100-fold more cold *H19* DMD4 competitor DNA was required to inhibit CTCF from binding to the C(−43)G mutant probe than to the wild-type *XIST* probe. These data indicate that the C(−43)G mutation results in a significantly enhanced binding of CTCF to the *XIST* minimal promoter.

These data were additionally supported by the results shown in Figure 4B. In that set of experiments, a radioactive high-affinity probe of the chicken FII insulator (24) was titrated with cold wild-type and C(−43)G mutant fragments. The wild-type fragment was not very efficient in competing with the FII fragment, whereas the C(−43)G mutant fragment completely abolished formation of hot probe–CTCF complexes, even at lowest concentration used. Titration with a cold FII fragment is shown as a control in Figure 2C.

Differences in the base-contacting points in CTCF–DNA complexes formed with the wild-type and mutated *XIST* promoters

To identify potential changes in contact guanine nucleotides recognized by CTCF in human wild-type and mutant *XIST*

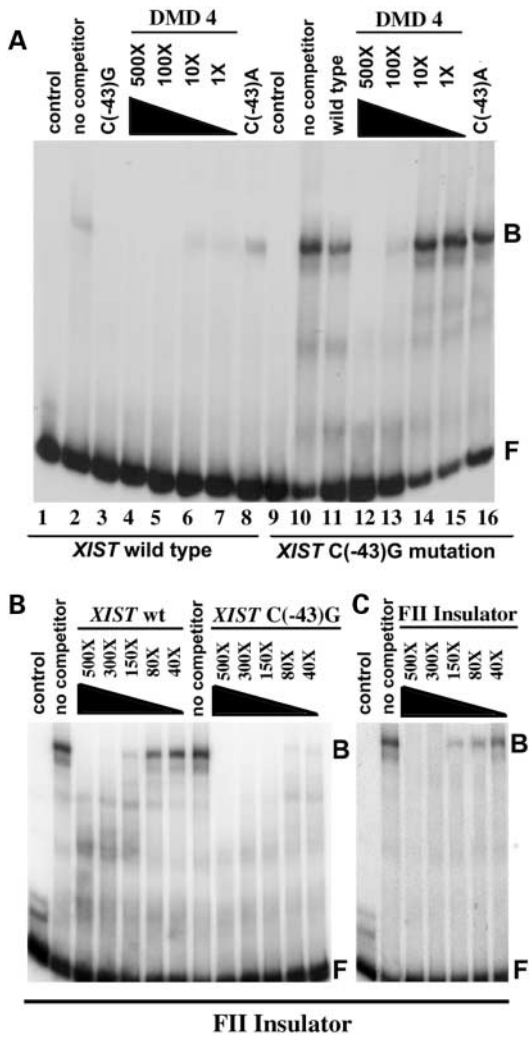

Figure 4. Competition EMSA experiments with wild-type, C(−43)G and C(−43)A mutant *XIST* minimal promoters. **(A)** EMSA reactions were carried out to compare relative binding of CTCF with the wild-type and with the C(−43)G-mutated *XIST* promoter regions in the presence or absence of cold (unlabeled) competitor DNA fragments. Lanes 1 and 9 (control) present EMSA with IVT luciferase. Lanes 2 and 10 (no competitor) present EMSA results with the recombinant CTCF in the absence of competitors. Lane 3 [C(−43)G] shows outcome from the competition by a 100-fold molar excess of unlabeled C(−43)G mutant probe for binding of CTCF to end-labeled wild-type *XIST* probe. Lane 11 (wild-type) shows outcome from the competition for binding of CTCF to the labeled C(−43)G-mutated probe with a 100-fold molar excess of unlabeled wild-type probe. Lanes 4, 5, 6 and 7 (500×, 100×, 10× and 1×, respectively) show results of the competition for binding of CTCF to the labeled wild-type *XIST* probe with a 500-, 100-, 10- and 1-fold molar excess of unlabeled DMD4 fragment (for the length and sequence of the DMD 4) (23). Results of the similar competition experiments but with the labeled C(−43)G mutant fragment used as EMSA probe are shown in lanes 12, 13, 14 and 15 (500×, 100×, 10× and 1×, respectively) with decreasing molar excess of the cold DMD4 fragment. Lanes 8 and 16 [C(−43)A] show a competition with 100-fold molar excess of unlabeled C(−43)A mutant probe for binding of CTCF to either wild-type or C(−43)G-mutated labeled probes, respectively. **(B)** Results of a similar competitive EMSA experiments but with the labeled chicken FII insulator probe in place of the hot *XIST* probes, and with unlabeled DNA fragments from either the wild-type or the C(−43)G-mutated *XIST* promoter regions. **(C)** Results of the control self-competition EMSA experiments with the chicken FII insulator probe. Naming of the EMSA gel lanes is similar to that in (A) and (B). On all panels, black triangles represent decreasing amounts of unlabeled competitor.

fragments, we carried out DMS-methylation interference assays for both strands of each DNA fragment. Guanine residues essential for the CTCF–DNA interaction are significantly reduced or excluded from the bound (B) DNA probe lane when compared with the unbound or free (F) probe lane (Fig. 5A). CTCF-contacting guanines for both wild-type and mutant probes are located within the DNA sequence from positions −44 to −36 upstream from the transcription start site (+1), revealing that the C(−43)G mutation hits the core of the CTCF-binding site. In comparing major CTCF-contacting guanines in the wild-type sequence and in the C(−43)G-mutated sequence, we noted that the C to G substitution creates a new contact nucleotide for CTCF binding in the top strand, and also results in elimination of one of the many dG-contacts in the bottom strand (Fig. 5A). Thus, this mutation alters the pattern of CTCF-contacting DNA bases, thereby yielding a new CTCF-binding sequence. Taken together with the observation that the *XIST* promoter fragment with the C(−43)A mutation did not bind CTCF, this result documents that the two types of germline mutations at position −43 of the *XIST* promoter are associated with drastic but opposite effects on CTCF–DNA interactions at this site.

To define which region of the EMSA positive fragment is covered by CTCF, we performed DNase I footprinting assays. Figure 5B shows that with the wild-type probe, CTCF protects a region of ∼30 bp from nuclease attack on the top strand and ∼40 bp on the bottom strand. Both top and bottom strands of the footprints showed several sites of accessibility to nuclease attack within the normal DNA–CTCF complex, which are not detected in the DNA–CTCF complex formed with DNA bearing the C(−43)G mutation. Nuclease hypersensitive sites, marked 'HS' in Figure 5B, are induced by CTCF binding at two positions inside and at the flanks of the footprints on both the wild-type and the mutant probes. CTCF protects more sequence on the top strand of DNA containing the C(−43)G mutation than is found with the wild-type *XIST* promoter region.

The observation that different amounts of CTCF protein were required for protection of wild-type and C(−43)G mutant promoters is in further keeping with increased CTCF binding to the C(−43)G mutant *XIST* promoter. Although one footprint unit of recombinant CTCF protein (see Materials and Methods for definition of the 'footprint unit') resulted in the complete protection of the C(−43)G mutant site from nuclease attack, the same amount of CTCF protein did not protect the wild-type XIST promoter region from the DNase I attack (data not shown). However, increase in CTCF to DNA ratio, as shown in Figure 5B, generated a visible DNase I footprint on the bottom and top strands of the wild-type *XIST* promoter region, but the pattern of protection was not as strong as with the C(−43)G mutant *XIST* promoter region, thereby suggesting a difference in CTCF–DNA complex formation and in affinity of the two sites. These results are summarized in Figure 5B–D.

Figure 5E represents the alignment of mouse and human *Xist* promoters in the region of the CTCF footprint. The alignment was obtained by the GCG-package Best Fit program for maximal nucleotide identity in the two sequences, with a 10 bp incremental extension in the 5′ direction starting from the first 10 bp window set for the +1 to +11 positions. The program generated this alignment for the two CTCF-binding DNA fragments from human and mouse promoters used in our EMSA analyses (Fig. 1). The alignment is displayed only for the top strands in the 5′ to 3′ direction corresponding to the 5′ flanking regions of *XIST/Xist* promoters immediately upstream of the transcriptional initiator (INR) start site conserved in human, mouse, rabbit and horse *Xist* promoters (17).

In contrast to the alignment of longer *Xist*-promoter sequences by Hendrich *et al.* (17), the alignment optimized for 84 bp demonstrates that all dG-contacts essential for CTCF binding on both strands in the human *XIST* promoter are conserved in the mouse *Xist* promoter, including the guanine at position −43 marked in red as a capital C in the top strand and underlined. Moreover, the nucleotide affected by the germline mutations belongs to the core of critical CTCF ZF-contacting bases of the wild-type human site, whereas the best fit alignment indicated that the same critical core motif of CTCF-contacting nucleotides is conserved in the wild-type mouse *(M. musculus) Xist* promoter. Note that both the number and arrangement of the aligned ZF-contacting bases depicted in capital letters in Figure 5E follow very well the stereo-chemical rules for DNA recognition by CTCF-like ZFs (25) provided that (i) approximately 4–5 nucleotides can indeed be engaged by one single finger of CTCF into binding to both mouse and human *Xist/XIST* sites, as was previously demonstrated for another CTCF site by analyzing the effects of CTCF ZF-deletions on the total length of DNase I CTCF-footprint in the APP-promoter (20), and (ii) a certain flexibility for having one or two non-contacting nucleotides between the groups of the dG-contacts made by the two neighboring ZFs is allowed by the presence and by the fitting positioning of non-canonical long inter-finger linker-regions in CTCF, as previously suggested by Kim and Pabo (26).

Reasoning that a different pattern of DNA protection by CTCF seen with wild-type and C(−43)G mutant promoter sequences might result from different CTCF conformations upon binding to the wild-type and to the mutant *XIST* DNA fragments, we next examined which particular sets of individual ZF from the 11 ZF DNA-binding domain are engaged in the binding of CTCF to these two target sequences.

Formation of CTCF–DNA complexes on the *XIST* promoter containing the wild-type sequence or the 'skewing mutation' in the −43 target site is mediated by dissimilar contributions of different ZF subsets

We have previously demonstrated that CTCF uses different sets of ZFs to bind different DNA sequences (15). Here, we performed EMSA of the human and mouse wild-type and human C(−43)G mutant *XIST* promoters with five N-terminally ZF truncated and six C-terminally ZF truncated *in vitro* translated products of the 11 ZF domain of CTCF, one-by-one from each side of this DNA-binding domain (27) (Fig. 6). The results are summarized in three panels (Fig. 6 A–C) with each of the 11 boxes representing one individual ZF for C- and N-terminal consecutive ZF deletions separately. A black box depicts that a particular ZF is dispensable for the binding to the particular site, a gray box indicates incomplete loss of binding and

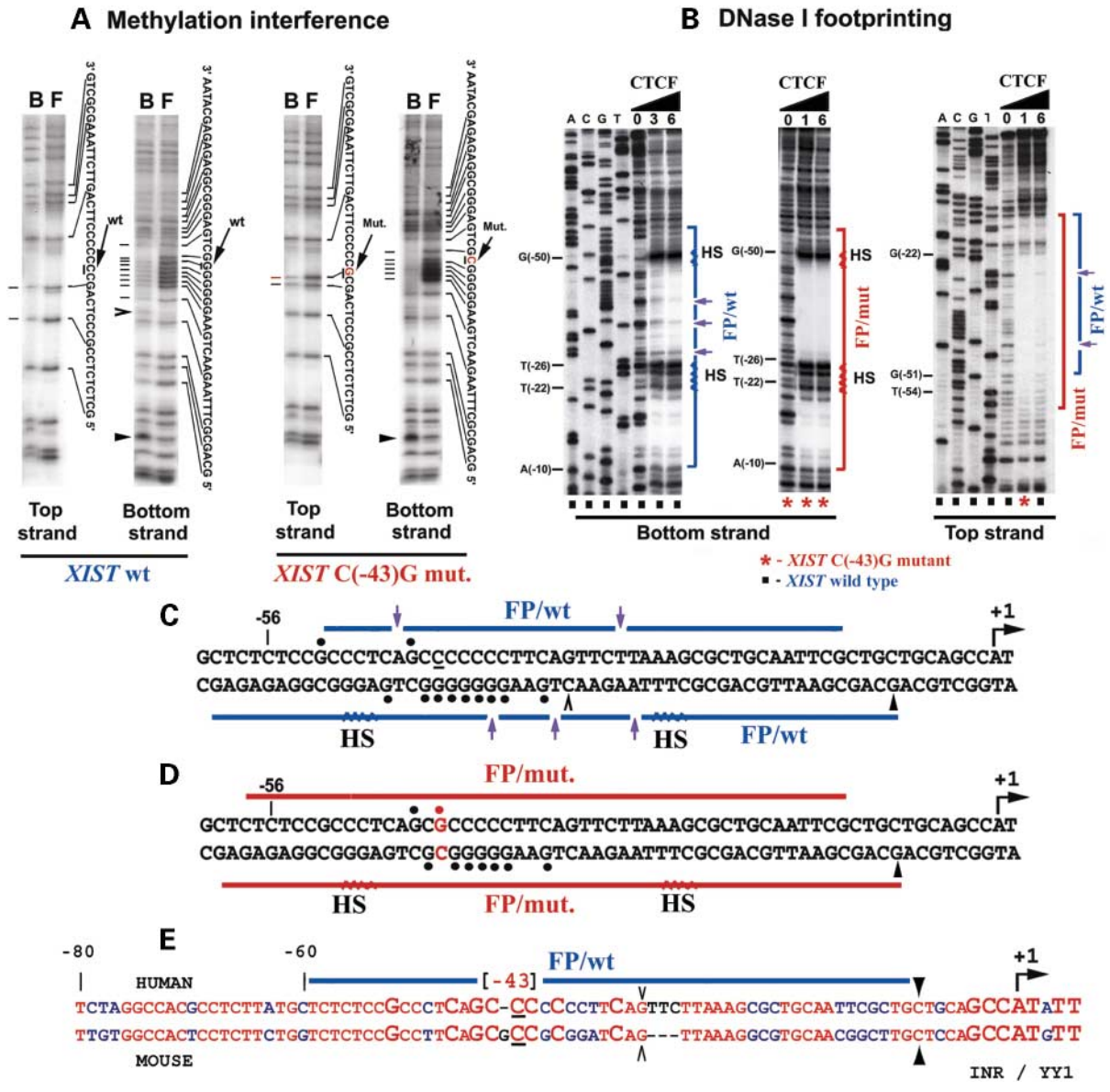

Figure 5. Characterization of CTCF–DNA complexes formed with the wild-type and C(−43)G mutant *XIST* promoter fragments by DNase I footprinting and DMS interference assays. **(A)** Methylation interference assays with the wild-type and the C(−43)G mutant *XIST* minimal promoter probes and the full-length CTCF. These assays identify individual DNA G-bases that cannot be modified by DMS without loosing or reducing CTCF binding. F, free (unbound) DNA probes, separated from the CTCF-bound (B) probes. DNA bases marked with bars are essential for CTCF binding. DMS-methylated dG^{N2}- residues preferentially associated with CTCF–DNA complexes are indicated by filled arrowheads. Open arrowhead points to the conformation-change in the bottom strand of the binding site. Wild-type C(−43)G-pair is indicated by an arrow named 'wt', whereas the mutant GC-pair is highlighted in red and indicated by an arrow named 'Mut.' **(B)** *In vitro* DNAse I footprinting of wild-type and C(−43)G mutant *XIST* minimal promoter fragment no. 8 with recombinant CTCF. All lines and letters marked in red color correspond to C(−43)G mutant *XIST* promoter, whereas those marked by a blue color correspond to wild-type *XIST* promoter. The 'A, C, G, T' depict the Maxam–Gilbert sequencing ladders. FP, the footprint regions protected from nuclease attack. HS, DNase I hypersensitive sites induced upon CTCF binding. The black triangle shows the increasing amount of footprint units of recombinant CTCF protein (see Materials and Methods) required for protection from nuclease attack. The numbers under triangle indicate the amount of footprint units added to the binding reaction. Results obtained with each strand of the C(−43)G-mutated probes are marked with red-colored asterisks, whereas black-filled squares show footprinting results on each strand of the wild-type probe. **(C)** Sequence of the wild-type *XIST* promoter with black circles showing contact nucleotides necessary for CTCF binding and with blue lines showing the regions protected from DNase I (FP). Violet vertical arrows point the sites of internal accessibility to nuclease inside of the wt DNA–CTCF complex. **(D)** Sequence of C(−43)G-mutated *XIST* promoter with black circles showing contact nucleotides for CTCF binding and with red lines showing the regions protected from DNase I (FP). The mutated contact nucleotide at −43 position is marked in red. The numbers indicate nucleotide positions relative to the +1 transcriptional start site of the *XIST* gene. **(E)** An optimal alignment of mouse and human *Xist* promoters in the region of CTCF footprint. The alignment is displayed only for the top strands in the 5′ to 3′ direction corresponding to the 5′-flanking regions of *XIST*/*Xist* promoters. The INR motif is shown in red capital letters. Because only the top strands are depicted, all CTCF-contacting Gs from the bottom strand of the DNase I footprint on the wt human site are emphasized as capital 'Cs'. Identical nucleotides are outlined in red font, whereas the differences are indicated in blue font. Four nucleotides, opposite to four single nucleotide gaps introduced to achieve the highest percent identity, are shown in black font, with the cytosine in the −43 position marked in red and underlined. No similarly significant homology was found immediately upstream of +100 position.

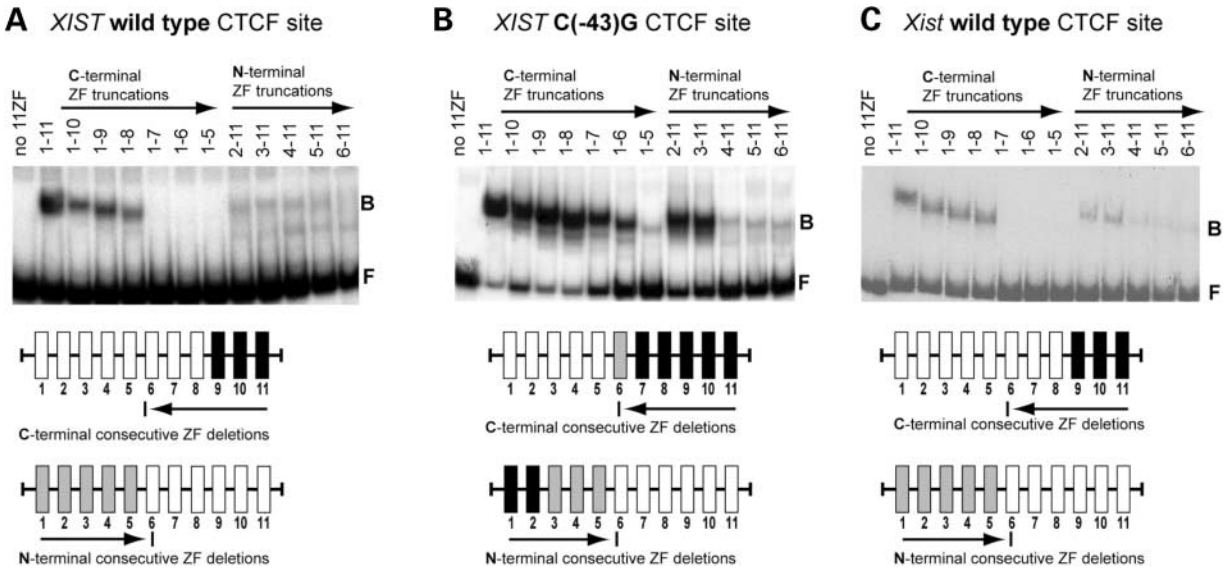

Figure 6. CTCF utilizes different combinations of ZFs for interaction with the wild-type and C(−43)G-mutant *XIST* minimal promoter. Two numbers that label each gel-lane indicate the first and the last ZF of each truncated form of the 11-finger CTCF DNA binding domain. Full-length 11-ZF polypeptide (lanes 1–11) and six C-terminal ZF-truncations (lanes 1–10 and 1–5) and five N- terminal ZF-truncations (lanes 2–11 and 6–11) were synthesized *in vitro* as described in Materials and Methods. Equal amounts of the CTCF proteins containing different groups of ZFs were analyzed by EMSA with human and mouse *XIST* wild-type, and with the C(−43)G-mutated probes. The positions of the unbound (F) and protein-bound (B) DNA probes are indicated. **(A–C)** The results of EMSA for each CTCF target are summarized in two 11 box panels to schematically depict either C-terminal or N-terminal consecutive ZF deletions in the upper and lower cartoons, respectively. Each box represents an individual ZF. Black boxes show that CTCF ZF can be deleted from the 11 ZF domain without significantly losing CTCF binding to the given DNA, gray boxes indicate incomplete loss of binding, whereas white boxes indicate ZFs that were essential for the CTCF–DNA interaction.

hence only modest contribution for binding, whereas white boxes indicate ZFs which are absolutely essential for the CTCF binding to each site.

The wild-type promoters of both humans and mice showed practically identical modes of ZF utilization for CTCF binding (Fig. 6). For instance, C-terminal finger 8 is required for binding to both human and mouse *XIST*, because deletion of this ZF completely abolished CTCF binding of constructs 1–7, 1–6 and 1–5 to the human and mouse wild-type promoters. In contrast, C-terminal fingers 8 and 7 are not important for binding to the promoter with the C(−43)G mutation, whereas deletion of ZF 6 only partially reduced CTCF binding to this fragment (Fig. 6). On the other hand, N-terminal fingers 1 and 2 are dispensable for CTCF binding to the promoter with the C(−43)G mutation, but they significantly reduced CTCF binding to the human and mouse wild-type promoters. We, therefore, conclude that although the pattern of ZF utilization has been stable since the divergence of a common ancestor between mouse and human, a single mutation at the *XIST* promoter associated with skewed X inactivation dramatically changes the ZF utilization. We propose that this change in pattern directly or indirectly leads to conformational changes of CTCF that may be of functional importance.

Increased CTCF affinity to the C(−43)G mutant *XIST* promoter is paralleled by gain of enhancer-blocking activity

We have earlier demonstrated that the binding of CTCF to a single target generally correlates with the enhancer-blocking activity of a target in an affinity-dependent manner (28). By extrapolation, changes in CTCF-binding efficiency induced by the −43 *XIST* site mutations identified *in vitro* by DNAse I footprinting and EMSA might be accompanied with similar changes in insulator properties. To this end, fragments (−308 to +31) with human and mouse wild-type, with C(−43)G and C(−43)A germline mutations and with the designer dG-contact mutations were inserted (in the same orientations) between the ß-globin LCR enhancer and the neomycin reporter gene (Fig. 7). Following transfection and neomycin selection, we scored for numbers of colonies. In this assay, the insulator strength is inversely proportional to the numbers of surviving cells that register as neomycin-resistant colonies (24). Figure 7 shows that mouse and human wild-type CTCF-binding region, or the C(−43)A and dG-contacts mutant *XIST* promoter fragments, do not noticeably reduce the number of colonies when compared with the transfection control used in the enhancer-blocking assay. In a sharp contrast, the same transfection-selection assay with the same plasmid except introducing of the C(−43)G mutation results in a significantly reduced colony numbers. This result strengthens the link between CTCF-binding efficiency and functional properties at the *XIST* promoter.

DISCUSSION

The convergence of C→A and C→G mutations at the same (−43) position upstream of the *XIST*+1 transcription start site and their association to opposite patterns of XCI (13,14) potentially provides an important opening with respect to

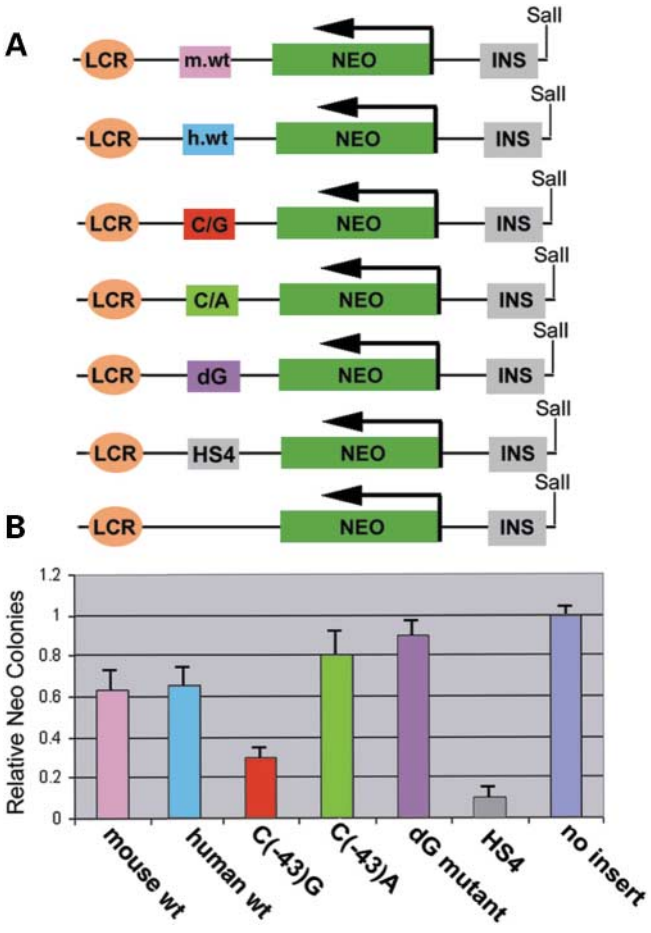

Figure 7. Enhancer-blocking activity of wild-type and mutant *XIST* promoter fragments. Human and mouse wild-type, C(−43)G, C(−43)A and dG-contacts mutant forms of the *XIST* 340 bp promoter fragments were inserted to the *Sac* I site positioned between the enhancer and Neo reporter gene in the pJC 5-4 construct. The original pJC 5-4 plasmid (24) was used as positive control for insulator activity. The pJC 5-4 vector without any insert at the *Sac* I site was used for normalization of background colony numbers. The error bars represent the standard deviation of three or four independent transfections with each vector.

our understanding of the X inactivation mechanism. Indeed, the identification of a factor that binds to this region, which is affected by these mutations, will provide an important handle to uncover this phenomenon. On the basis of ChIP, EMSA, DNase I footprinting and methylation interference assays, we argue here that this factor is identical to CTCF. Not only do these mutations map to an ∼40 bp long CTCF-binding sequence within a string of guanines at −44 to −36 positions involved in establishing the DNA–protein contacts necessary for the specific interaction with CTCF, but CTCF also interacts with the *Xist* promoter in only female human and mouse cells. This could be explained by the demonstration here that CTCF associates to the preferentially active *Xist* allele in female mouse placenta. Taken together with a cluster of CTCF-binding sites in the chromatin insulator region earlier mapped at the 3′ end of *Xist* (11), these findings suggest a dual function for CTCF in the X inactivation process.

The changes in the affinity of CTCF for the wild-type and mutant *XIST* promoters might be explained by our observations of how distinctively each of the two mutations affected critical CTCF-contacting DNA bases. The C(−43)A mutation removed one base from the recognition site, whereas the C(−43)G mutation altered two CTCF-contacting bases by creating a new dG-contact in the top strand and removing another dG-contact from the bottom strand. The competitive EMSA titration experiments showed that the newly created *XIST* site with the C(−43)G mutation has a higher affinity for CTCF than two previously characterized insulator sites with extremely strong CTCF binding. One explanation to account for this effect is based on the observation that enhanced affinity of artificial poly-ZF proteins could be achieved by adjusting and selecting slightly longer linkers between six CTCF-like C2H2-class fingers (26). Because CTCF contains several similarly longer and non-canonical linkers, these may allow CTCF to bind to different DNA sequences in distinctly different ways (15) providing necessary flexibility to make different number of contacts, and to form structurally distinct complexes with varying stability. Moreover, the combination of ZF used for binding the wild-type and C(−43)G mutant *XIST* promoters differed rather significantly. Therefore, it is most likely that the interaction with the C(−43)G mutant site involves different ZF subsets for reaching a near-ideal 'fitting' of the new dG-contacts to the stereo-chemical rules (25).

Although neither the wild-type nor the C(−43)A mutant *XIST* promoter displayed any significant chromatin insulator function, the C(−43)G mutant *XIST* promoter was very potent in preventing enhancer–promoter communications. Because the chromatin insulator function could be assigned to the mutant *XIST* promoter fragment that is associated solely with the inactive X chromosome and hence an active *XIST*, it is possible that the insulator protects from repressive, upstream *cis* elements in the affected patients (17). Alternatively, and not mutually exclusive, the CTCF target site in the *XIST* promoter might facilitate the recruitment of the pre-initiation RNA polymerase complex to enhance *XIST* transcription. This possibility is supported by the demonstration that CTCF is a component of the RNA polymerase II holoenzyme complex and that a range of intergenic CTCF-binding sites throughout the mouse genome display extensive binding to pol II, as determined by double ChIP on chip assays (Chernukhin *et al.*, unpublished data). However, as indicated by any direct effect of the mutations on *XIST* promoter-directed transcription using transient transfection assays of plasmid DNA (data not shown), this possibility might apply only in the context of higher order chromatin conformations.

The skewed X inactivation profiles in patients with the point mutations at the pivotal −43 position of the *XIST* promoter suggest an interference with the random choice of XCI in affected patients. By extrapolation, the choice mechanism in normal somatic cells might be governed by the regulated affinity between CTCF and the wild-type *XIST* promoter. The *Tsix* transcript itself may be an important component of this process, because the *Tsix* transcriptional process which extends beyond the *Xist* promoter (8) could interfere with CTCF–*Xist* promoter interactions on both X chromosomes in females prior to the onset of the inactivation process.

Another parameter in this process might be the cooperativity of CTCF target sites in this region, as CTCF target sites within the *H19* imprinting control region have been shown to interact in a cooperative manner via heterodimerization (18). If so, the choice process could involve the transformation of an initially weak CTCF target site into a strong site by cooperative interactions between CTCF in *Xist* and neighboring CTCF target sites at the choice/imprinting center in the *Tsix* gene (11). This possibility is in keeping with the demonstration here that CTCF is stably associated with the *Xist* promoter of only the inactive X chromosome and that CTCF has an essential role in mediating higher order chromatin conformations. The CTCF-dependent close physical proximity between the *H19* ICR and a differentially methylated region (DMR)1 might be essential for the silencing of the maternal *Igf2* gene as well as the coordination of regional epigenetic marks (Tiwari *et al.*, submitted for publication). The existence of CTCF target sites within the DMR1 (Tiwari *et al.*, submitted for publication) might enable formation of the *H19* ICR–DMR1 complex via CTCF heterodimers. This proposal is in keeping with our demonstration that heterodimeric CTCF–DNA complexes organize interactions within the *H19* ICR (18). By extrapolation, CTCF might be involved in the choice mechanism by generating one kind of chromatin conformation on the active X chromosome from *Tsix* and another on the inactive X chromosome from *Xist.*

Another possibility for the choice mechanism is involvement of testis-specific CTCF-paralog BORIS (29). It shares the same 11 ZF-DNA recognition domain with CTCF, but has absolutely no homologies to CTCF in ZF-flanking N- and C-terminal domains. Male germ cells deficient in both CTCF and me-dC content and undergoing resetting of epigenetic marks at *H19* ICR CTCF sites (18,23,29) are positive for BORIS. Moreover, BORIS is active in undifferentiated but not in differentiated ES cells (data not shown). Taken together, these observations raise an interesting possibility that BORIS may participate in the choice mechanism for XCI, being expressed at the 'right place at right time'.

In conclusion, our results indicate a close association between the choice of the future inactive X chromosome and binding properties of a CTCF target site in the *XIST* promoter. Regardless of the underlying mechanisms, it is remarkable that a point mutation at a single but pivotal CTCF target site on one of the X chromosomes can be linked to the activation or inactivation of the opposite chromosome in humans. We postulate here that this process requires the CTCF-dependent organization of higher order chromatin conformation that impinges on the stability of the CTCF–*XIST* promoter complex during early stages of X inactivation.

MATERIALS AND METHODS

Tissue and cell materials

Fetal mouse placenta derived from *M.m. musculus* × *M. spretus* interspecific crosses was dissected under microscope and tissue cells dispersed prior subsequent analyses. The male and female fetal placentas (E14.5) were discriminated by PCR-based genotyping. Live primary cultures (at passages 2–4) of HMEC and NHDF derived from healthy adult men and women, purchased from Clonetics and Cambrtex, were used in ChIP assay to show an association between CTCF and the human *XIST* promoter. In addition, we used freshly isolated PBLs from normal adult XX versus XY blood. Blood (100 ml) was drawn by authorized personnel at the Red Cross Donor Facility (Rockville, MD, USA) from healthy male and female volunteers of 30–40 years of age, and PBL fraction purified using LSM (Mediatech Cellgro, VA, USA) according to the manufacturer's instructions. Approximately 100 million cells from isolated male and female PBL suspensions were taken for each preparation of fixed chromatin for each ChIP. Minor modifications for ChIP analyses with human cells included addition to all buffers immediately before use of protease inhibitors as recommended by a manufacturer (Sigma; Cat. no. P-8340).

Chromatin immunopurification analyses

DNA and protein components of mouse or human chromatin were cross-linked *in situ* by incubation of dispersed cells in 1% formaldehyde for 10 min at 37°C. The DNA–protein complexes were immunopurified using anti-CTCF antibody (Upstate) or a mixture of nine anti-CTCF MABs (Supplementary Material) and protein A4 Fast Flow Sepharose beads (Pharmacia-Upjohn) as described earlier (23).

The immunopurified DNA from mouse tissues was PCR amplified with primers spanning the *Xist* promoter: sense primer 5′-CACGCGTCATGTCACTGAGCTTACGTACCTC-3′; antisense primer 5′-GAACCGCACATCCACGGGAAAC GAG-3′. PCR conditions were 95°C for 5 min, 29×(94°C for 45 s, 62°C for 1 min, 72°C for 45 s). As an internal control to allow comparison between the formaldehyde cross-linked samples, the CTCF target site no. 3 within the *H19* ICR was amplified using primers: sense primer 5′-CTCAGTGGTCGA TATATGGTTT-3′; antisense primer 5′-TGAGTCAAGTTCT CTTGGTTC-3′. PCR conditions were 95°C for 3 min, 28×(94°C for 40 s, 54°C for 40 s, 72°C for 40 s). The PCR products were analyzed on a 1.5% agarose gel and visualized by SYBR® Green. To distinguish the maternally inherited allele from the paternally inherited one in mouse placentas from female progeny of two different strains, the CTCF chromatin immunopurification (ChIP) material from the interspecific cross was subjected to sequencing exploiting an A/C polymorphism at position −88.

The immunopurified DNA of human origin was PCR amplified with primers spanning the *XIST* promoter: sense primer 5′-caaatcacaaagatgtccggctttcaatcttctaggc-3′ and antisense primer 5′-aagcttccagccccgagagagtaagaaatatggctg-3′. As positive controls for the human *XIST* site in CTCF ChIP analyses, we used well known human *c-MYC* oncogene 5′-insulator site (N-site) that was amplified by using primers: sense primer 5′-cctgaaagaataacaaggaggtggctggaaacttg-3′and antisense primer 5′-gcaaattactcctgcctccaggcctttg-3′.

Quantitation of ChIP DNA was performed by real-time PCR analysis using the ABI Prism 7900 Sequence Detection System according to Applied Biosystem's SYBR® Green PCR Master Mix Protocol. Real-time PCR was carried out in triplicate of ChIP, control and input DNA samples at the following thermal cycling parameters 95°C for 10 min and

40 cycles of 95°C for 15 s and 60°C for 1 min. Data were collected at 60°C and analyzed by applying comparative C_T method as described in ABI User Bulletin (Biosystems, updated 04/2001 #933) and in Litt *et al.* (30). Fold enrichment for a particular target sequence was determined by calculation of ratio of the amount of the target sequence in immunoprecipitation (IP) to the amount of the target sequence in input (In) DNA. Briefly, we used the equation

$$\frac{X_0(\mathrm{IP})}{X_0(\mathrm{In})} = 2^{C_T(\mathrm{IP})-C_T(\mathrm{In})}$$

where X_0 is the initial DNA concentration of a target sequence in IP and In and C_T is a number of cycles required to reach final concentration of target DNA X_n, which is inversely related to the initial target sequence concentration X_o and determined by using ABI SDS 2.1 software.

Sequencing of mouse *Xist* promoter

To find a sequence polymorphism(s) at or near the CTCF-binding site, we analyzed the DNA sequence of mouse *Xist* minimal promoter from five strains of mice [129X1/SvJ mouse_(Xce-a); C57BL/6J mouse_(Xce-a); BALB/cJ mouse_(Xce-b); CAST/EiJ mouse_(Xce-c); and SPRET/Ei mouse_(Xce-d)] with the previously characterized strain-specific haplotypes of the 'X-controlling element' (Xce) from isolated DNA purchased from the Jackson Laboratory. Mouse DNA from all mouse strains was amplified from position −32 to +690 relative to the *Xist* transcriptional start site by sense primer: 5′-GACAGTTCTTTTAAGTTAGCAGTGTCTCTGG GG-3′ and antisense primer 5′-GAACCGCACATCCACGGG AAACGAGC-3′. The amplification protocol was as follows: denaturation at 95°C for 5 min, followed by 35 cycles at 95°C for 1 min, 66°C for 1 min, 72°C for 1 min and final extension 72°C for 10 min. The primers used for PCR amplifications were designed based on the high-quality wild-type *XIC*/*Xic* sequence annotations (31). The nucleotide sequences were submitted to GenBank with accession nos AY618353 for 129X1/SvJ mouse_(Xce-a); AY618354 for C57BL/6J mouse_(Xce-a); AY618355 for BALB/cJ mouse_(Xce-b); AY618356 for CAST/EiJ mouse_(Xce-c); AY618357 for SPRET/Ei mouse_(Xce-d).

Plasmid constructs

The PCR fragment of the human, CTCF-positive *XIST* minimal promoter was cloned into the pCRII-TOPO vector (Invitrogene) resulting in the wt*XIST*-pCRII-TOPO plasmid. The PCR fragment was amplified from position −309 to +31 relative to the *XIST* transcriptional start site by sense primer: 5′-GTTATGGAGGATTTTAGCATTAATTATTG-3′ and antisense primer 5′-CCAGCCCCGAGAGAGTAAGAA TATG-3′. The *Sac* I sites at the 5′ end of primers were introduced to facilitate the subcloning of the fragment into either the enhancer-blocking plasmid or the luciferase-reporter pGL-2 plasmid. The amplification protocol was as follows: denaturation at 95°C, 5 min, followed by 35 cycles at 95°C for 30 s, 64°C for 30 s, 72°C for 1 min, final extension 72°C for 10 min.

Both the naturally occurring C(−43)G and C(−43)A mutations and the artificial GCCCCCC/AAAGCTT mutation (named dG-contacts mutation) were generated in the wt*XIST*-pCRII-TOPO plasmid by the QuikChange site-directed mutagenesis kit (Stratagene). DNA sequencing on Li-Cor Sequencer confirmed the accuracy of the cloning as well as the mutagenesis of the insert. These plasmids [wild-type, the C(−43)G and C(−43)A mutations and the dG-contacts mutation] from the *XIST*-pCRII-TOPO-based vectors were used for generating of insulator constructs, and as the templates for DNase I footprinting and methylation interference analyses.

The insulator assay series of plasmids were generated on the basis of pJC 5-4 plasmid as has been described earlier (32). The HS4 globin insulator was inserted into *Sac* I site of the parent plasmid between an enhancer and the mouse γ-globin promoter–Neo reporter gene. To substitute this *Sac* I HS4 insulator with the wild-type, C(−43)G mutant, C(−43)A mutant and 'dG' mutant *XIST* promoter fragments, the inserts were cut out by *Sac* I enzyme from the corresponding pCRII-TOPO-based constructs described earlier, and ligated into the *Sac* I site of the pJC 5-4 plasmid. To normalize the colony number, we used pJC 5-4 vector with deleted HS4 globin insulator.

Nuclear extracts and *in vitro* transcription–translation

Full-length human CTCF and the 11 ZF CTCF-binding domain were *in vitro* translated from pET-7.1 and pET-11 ZF, respectively (33), using the TnT reticulocyte lysate-coupled *in vitro* transcription–translation system (Promega). Twelve plasmids with sequentially truncated ZFs of the CTCF-binding domain (from amino acid position 236–622 of the human CTCF protein) were *in vitro* synthesized in the presence of ^{35}S-Met (27). Nuclear extracts were prepared from K562 and other cell lines by using the 2 × NUN extraction solution as described elsewhere (34). Partially purified recombinant CTCF was obtained from *Pichia pastoris* extract by single step SP cation exchange chromatography as described previously (20). Fractions of 0.5 ml were collected and CTCF-binding activity was monitored by EMSA with CTCF target from the APP promoter (20). The fraction containing peak binding activity was used for DNase I footprinting to determine the minimal amount (1 μl) of recombinant CTCF protein sufficient to fully protect CTCF-binding site in the APP promoter from the nuclease attack. In the experiments with the *XIST* promoter, 1 μl of this fraction was taken as 'one CTCF footprint unit'.

Electrophoretic mobility shift assays

Eight consecutive fragments of the human *XIST* minimal promoter from position −1080 to +31 relative to +1 start site and eight consecutive fragments of the mouse *Xist* minimal promoter from position −1020 to +34, one-by-one covering partially overlapping promoter DNA sequences of interest, were PCR amplified. The fragments were simultaneously end-labeled on either strand during amplification by using pairs of 20–28 bp long primers that were end-labeled at their 5′ ends using ^{32}P-γ-ATP and T4 polynucleotide kinase.

The DNA–protein binding incubation was carried out in a buffer containing standard phosphate-buffered saline with 5 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 1 mM dithiothreitol, 0.1% Nonidet P-40 and 10% glycerol in the presence of poly (dI–dC) plus poly (dG)·poly (dC). The nuclear extract-based EMSA was performed in the presence of unlabeled double-stranded oligonucleotides containing Sp1 like sites and 'G-string'-binding factors as described (35). To reach the same shift mobility for *in vitro* translated CTCF and nuclear extract's CTCF, we added 10 μl of a TnT reticulocyte lysate to the nuclear extract-based binding reaction. For 'super-shift' assays, an antibody against the C-terminal domain of CTCF (Upstate Biotechnology) was used. The EMSA reaction mixtures of a 20 μl final volume were incubated for 30 min at room temperature followed by electrophoresis on 5% non-denaturing polyacrylamide gels.

The competition assay was performed with wild-type, C(−43)G and C(−43)A mutant *XIST* fragments PCR amplified by sense primer: 5′-CAAATCACAAAGATGTCCGGC TTTC-3′ and antisense primer 5′-CCAGCCCCGAGAGAG TAAGAATATG-3′ from wt *XIST*-, C/G- and C/A-mut.*Xist*-pCRII-TOPO plasmids, respectively. The specificity of the protein–DNA interactions were assessed by adding 100-fold molar excess of unlabeled wild-type or C/G mutant probes in reactions with full-length CTCF and end-labeling C/G mutant or wild-type PCR fragments, respectively. The C(−43)A mutant was used as cold competitor in reaction with full-length CTCF and end-labeling wild-type and C/G mutant PCR fragments in 100-fold molar excess. Finally, the well-characterized CTCF-binding site, site no. 3 from the human *H19* imprinting control region (23), was used with a 500-, 100-, 10- and 1-fold molar excess of cold probe.

DNase I footprinting and methylation interference assay

We performed DNase I footprinting and methylation interference analyses with *in vitro* translated, full-length CTCF and the 141 bp *XIST* wild-type and C(−43)G mutant promoter fragments amplified by sense primer: 5′-CAAATCACAAA GATGTCCGGCTTTC-3′ and antisense primer 5′-CCAG CCCCGAGAGAGTAAGAATATG-3′ from wt*XIST*- and C/G-mut.*XIST*-pCRII-TOPO plasmids, respectively. Both fragments were labeled at their 5′ ends on either the top or the bottom strand. Following gel purification, the fragments were either incubated with CTCF and then partially digested with DNase I or partially methylated at guanine residues by dimethyl sulfate and then incubated with CTCF, as has been described (27).

Insulator assay

Insulator assays were performed in human K562 erythroleukemia cells as has been described (24). All plasmid DNAs, prepared by using Qiagen Maxiprep Kit, were linearized with *Sal* I, phenol–chloroform extracted and ethanol-precipitated. In total, 100 ng of linear plasmid DNA was mixed with 10^7 of K562 erythroleukemia cells and electroporated at a pulse field strength of 200 V and 960 μF capacitance with no resistor. The transfected cells were grown in 20 ml DMEM medium containing 10% fetal bovine serum without G418 at 37°C for 24 h. The following day, the cells were seeded on 150 mm tissue culture plates in soft agar prepared by standard procedure with DMEM + FBS containing 750 μg/ml of G418. Finally, the numbers of G418-resistant cell colonies formed during 3 weeks of neomycin selection were counted. In this assay, which was repeated at least three times, the relative numbers of colonies reflect the potency of an insert to block enhancer-mediated activation of the Neo-driving promoter.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We wholeheartedly thank Dr Herbert Morse III for critical reading of the manuscript, Dr Anita Mattsson for the mouse placenta work, Drs Chen-Feng Qi and Shao Xiang for performing the immunoblot analysis of nine CTCF monoclonal antibodies, and Drs Yoo-Wook Kwon and Hanlim Moon and other members of the Lobanenkov Lab for encouragement and support during the course of this work. We would also like to acknowledge an anonymous reviewer suggested our search for sequence polymorphisms near CTCF site by sequencing *Xist* promoters from five particular mouse strains. This work was supported by intramural research funding to V.V.L. and grants from the National Institutes of Health (Grant NS30994) to W.Q., the Swedish Science Research Council (VR) to R.O., the Juvenile Diabetes Research Foundation International (JDRF) to R.O., the Swedish Cancer Research Foundation (CF) to R.O., the Swedish Pediatric Cancer Foundation (BCF to R.O. and the Wallenberg and Lundberg Foundations to R.O.

REFERENCES

1. Lyon, M.F. (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Naturwissenschaften*, **190**, 372–373.
2. Lyon, M. (1999) Imprinting and X chromosome inactivation. In Ohlsson, R. (ed.), *Genomic Imprinting: An Interdisciplinary Approach*. Springer-Verlag, Berlin/Heidelberg/New York, Vol. 25, pp. 73–90.
3. Takagi, N. and Sasaki, M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*, **256**, 640–642.
4. West, J.D., Frels, W.I., Chapman, V.M. and Papaioannou, V.E. (1977) Preferential expression of the maternally derived X chromosome in the mouse yolk sac. *Cell*, **12**, 873–882.
5. Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S. and Brockdorff, N. (1996) Requirement for Xist in X chromosome inactivation. *Nature*, **379**, 131–137.
6. Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J. and Willard, H.F. (1992) The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, **71**, 527–542.
7. Clemson, C.M., McNeil, J.A., Willard, H.F. and Lawrence, J.B. (1996) XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.*, **132**, 259–275.
8. Lee, J.T., Davidow, L.S. and Warshawsky, D. (1999) *Tsix*, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.*, **21**, 400–404.
9. Stavropoulos, N., Lu, N. and Lee, J.T. (2001) A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc. Natl Acad. Sci. USA*, **98**, 10232–10237.

10. Ogawa, Y. and Lee, J.T. (2003) Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol. Cell*, **11**, 731–743.
11. Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S. and Lee, J.T. (2002) CTCF, a candidate *trans*-acting factor for X-inactivation choice. *Science*, **295**, 345–347.
12. Nesterova, T.B., Johnston, C.M., Appanah, R., Newall, A.E., Godwin, J., Alexiou, M. and Brockdorff, N. (2003) Skewing X chromosome choice by modulating sense transcription across the Xist locus. *Genes Dev.*, **17**, 2177–2190.
13. Plenge, R.M., Hendrich, B.D., Schwartz, C., Arena, J.F., Naumova, A., Sapienza, C., Winter, R.M. and Willard, H.F. (1997) A promoter mutation in the *XIST* gene in two unrelated families with skewed X-chromosome inactivation. *Nat. Genet.*, **17**, 353–356.
14. Tomkins, D.J., McDonald, H.L., Farrell, S.A. and Brown, C.J. (2002) Lack of expression of XIST from a small ring X chromosome containing the XIST locus in a girl with short stature, facial dysmorphism and developmental delay. *Eur. J. Hum. Genet.*, **10**, 44–51.
15. Ohlsson, R., Renkawitz, R. and Lobanenkov, V. (2001) CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.*, **17**, 520–527.
16. Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R. *et al.* (2005) CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep.*, **6(2)**, 165–170.
17. Hendrich, B.D., Plenge, R.M. and Willard, H.F. (1997) Identification and characterization of the human *XIST* gene promoter: implications for models of X chromosome inactivation. *Nucleic Acids Res.*, **25**, 2661–2671.
18. Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., Klenova, E., Lobanenkov, V. and Ohlsson, R. (2004) Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of *de novo* methylation upon maternal inheritance. *Mol. Cell. Biol.*, **24**, 3497–3504.
19. Klenova, E.M., Morse, H.C., III, Ohlsson, R. and Lobanenkov, V.V. (2002) The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Semin. Cancer Biol.*, **12**, 399–414.
20. Quitschke, W.W., Taheny, M.J., Fochtmann, L.J. and Vostrov, A.A. (2000) Differential effect of zinc finger deletions on the binding of CTCF to the promoter of the amyloid precursor protein gene. *Nucleic Acids Res.*, **28**, 3370–3378.
21. Blanchette, M., Schwikowski, B. and Tompa, M. (2002) Algorithms for phylogenetic footprinting. *J. Comput. Biol.*, **9**, 211–223.
22. Garrett, F.E., Emelyanov, A.V., Sepulveda, M.A., Flanagan, P., Volpi, S., Li, F., Loukinov, D., Eckhardt, L.A., Lobanenkov, V.V. and Birshtein, B.K. (2005) Chromatin architecture near a potential 3′ end of the Igh locus involves modular regulation of histone modifications during B-cell development and *in vivo* occupancy at CTCF sites. *Mol. Cell. Biol.*, **25**, 1511–1525.
23. Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C.F., Wolffe, A., Ohlsson, R. and Lobanenkov, V.V. (2000) Functional association of CTCF with the insulator upstream of the *H19* gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.*, **10**, 853–856.
24. Bell, A.C., West, A.G. and Felsenfeld, G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, **98**, 387–396.
25. Suzuki, M., Gerstein, M. and Yagi, N. (1994) Stereochemical basis of DNA recognition by Zn fingers. *Nucleic Acids Res.*, **22**, 3397–3405.
26. Kim, J.S. and Pabo, C.O. (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl Acad. Sci. USA*, **95**, 2812–2817.
27. Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J. and Lobanenkov, V.V. (1996) An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian *c-myc* oncogenes. *Mol. Cell. Biol.*, **16**, 2802–2813.
28. Mukhopadhyay, R., Yu, W., Whitehead, J., Xu, J., Lezcano, M., Pack, S., Kanduri, C., Kanduri, M., Ginjala, V., Vostrov, A. *et al.* (2004) The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res.*, **14(8)**, 1594–1602.
29. Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A. *et al.* (2002) BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl Acad. Sci. USA*, **99**, 6806–6811.
30. Litt, M.D., Simpson, M., Recillas-Targa, F., Prioleau, M.N. and Felsenfeld, G. (2001) Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *EMBO J.*, **20**, 2224–2235.
31. Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P. and Duret, L. (2002) Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res.*, **12**, 894–908.
32. Chung, J.H., Bell, A.C. and Felsenfeld, G. (1997) Characterization of the chicken beta-globin insulator. *Proc. Natl Acad. Sci. USA*, **94**, 575–580.
33. Awad, T.A., Bigler, J., Ulmer, J.E., Hu, Y.J., Moore, J.M., Lutz, M., Neiman, P.E., Collins, S.J., Renkawitz, R., Lobanenkov, V.V. *et al.* (1999) Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence. *J. Biol. Chem.*, **274**, 27092–27098.
34. Lavery, D.J. and Schibler, U. (1993) Circadian transcription of the cholesterol 7 alpha hydroxylase gene may involve the liver-enriched bZIP protein DBP. *Genes Dev.*, **7**, 1871–1884.
35. Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V. and Goodwin, G.H. (1990) A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken *c-myc* gene. *Oncogene*, **5**, 1743–1753.
36. Heard, E. (2004) Recent advances in X-chromosome inactivation. *Curr. Opin. Cell Biol.*, **16**, 247–255.
37. Orlando, V., Strutt, H. and Paro, R. (1997) Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods*, **11**, 205–214.
38. Kuo, M.H. and Allis, C.D. (1999) *In vivo* cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods*, **19**, 425–433.