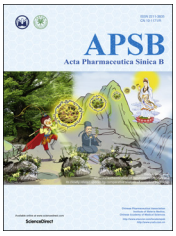
Acta Pharmaceutica Sinica B 2018;8(6):900–908

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



ORIGINAL ARTICLE

DNA recognition patterns of the multi-zinc-finger protein CTCF: a mutagenesis study

Jingjing Guo, Ni Li, Jiexiong Han, Fei Pei, Tianyu Wang, Duo Lu*, Jiandong Jiang*

State Key Laboratory of Bioactive Substance and Function of Natural Medicine, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China

Received 21 March 2018; received in revised form 4 June 2018; accepted 28 June 2018

KEY WORDS

CTCF;
Zinc-finger;
Structure integrity;
Mutagenesis;
DNA recognition patterns

Abstract CCCTC-binding factor (CTCF) is a zinc-finger protein, serving an important part in the genome architecture as well as some biochemical processes. Over 70,000 CTCF binding DNA sites have been detected genome-wide, and most anchors of chromatin loops are demarcated with the CTCF binding. Various protein or RNA molecules interact with DNA-bound CTCF to conduct different biological functions, and potentially the interfaces between CTCF and its cofactors can be targets for drug development. Here we identify the effective region of CTCF in DNA recognition, which defines the exposed CTCF surface feature for the interaction of cofactors. While the zinc-finger region contributes the most in DNA association, its binding affinity varies based on different DNA sequences. To investigate the effectiveness of individual zinc-fingers, the key residues are mutated to inactivate the DNA binding ability, while the finger configuration and the spacing between fingers are preserved. The strategy is proved to be successful, while clear differences are observed in the DNA binding affinities among the 11 finger mutants and the result is consistent to previous studies in general. With the help of inactivated finger mutants, we identify the ineffective fingers and the dominant effective fingers, which form distinctive patterns on different DNA targets.

*Corresponding authors. Tel.: +86 10 63039979; fax: +86 10 63017757.
E-mail addresses: luduo@imm.ac.cn (Duo Lu), jiangjd@imm.ac.cn (Jiandong Jiang).

Peer review under responsibility of Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association.

https://doi.org/10.1016/j.apsb.2018.08.007

1. Introduction

CCCTC-binding factor (CTCF) is a multivalent protein with various functions, and has drawn most attention for its involvement in the genome organization. Accumulating evidences suggest that a genome is organized into relatively insulated compartments, which are now called topologically associated domains (TADs)[1–3]. To fold into TADs, chromosomes take the conformation of intricate chromatin loops whose anchors are frequently found at the boundary of TAD and are bound with CTCF[3]. Mutations or modifications on CTCF or its DNA binding sites have been shown to cause TAD and loop structural alterations, which deregulated the enhancer–promoter communications of context genes in the surrounding TADs and led to various pathological consequences[4–6]. In addition to the function at loop anchors, which account for a portion of CTCF binding sites genome-wide, other functions have been reported for DNA-bound CTCFs such as stalling RNA polymerase II process and consequently affecting alternative splicing[7]. CTCF conducts its biological functions in association with many other protein factors, such as COHESIN, PolII, YY1, OCT4, and TAF3, through protein–protein interactions[8–10]. The interactions with the cofactors have been found to involve different parts of CTCF, which are mostly located at DNA-unbound regions.

CTCF was originally named for the presence of the repeating CCCTC sequence in its DNA binding site at the chicken *C-MYC* 5′-flanking sequence[11], and later studies revealed a consensus sequence of (T/C)GCCC(T/C)CTX(G/C)TGG at this core motif for CTCF recognition. Further analysis showed additional 2 motifs flanking the CCCTC motif, although they did not always present in a CTCF binding site[12]. Within the CTCF, there were 11 tandem adjacent zinc-finger motifs proposed to interact with DNA, while the flanking N and C terminal regions both accommodated some basic-residue rich sequences. The 11 zinc-fingers of CTCF were composed of 10 C2H2 fingers and the 11th C2HC finger (Fig. 1), which reassembled the C2H2 finger in structure.

Zinc-finger has been frequently found as a DNA binding motif in many transcription factors, while RNA and protein associating zinc-fingers have been reported as well[13,14]. The structural motif is composed of a short beta hairpin and an alpha helix held together by a zinc ion (Fig. 1). Based on the detailed studies of the zinc-finger DNA recognition, a general rule has been concluded: the residues at zinc-finger helical positions −1, 2, 3, and 6 make the specific interactions to the DNA bases located inside the major groove of double helix. Due to the nature of tandem connection between individual finger motifs, a set of zinc-fingers are capable of recognizing a continuous DNA sequence, while each finger covers 3–4 DNA base-pairs[15–17].

However, the number of tandem DNA recognizing zinc-fingers seldom exceeds three. This has been thought a consequence of periodic mismatch between the DNA double helix structure and a pack of zinc-fingers with standard linkers, generally TGEKP in sequence, which restrains extra zinc-fingers from making canonical contact with DNA bases[17]. Extended linker sequence or an unbound finger motif would allow further recognition on a DNA sequence separated from the existing site[18,19]. For CTCF, 7 out of its 10 linkers are in standard length (Fig. 1). The 6th and 7th linkers have one extra residue, and the 10th linker has 3 extra residues.

CTCF zinc-finger recognition modes were examined at several DNA sites with finger-truncated mutant proteins, which revealed the DNA binding effective regions comprising different sets of consecutive fingers[20–22]. The contribution of the internal fingers in the effective region was nevertheless unknown due to the restriction of the method. Another approach with structurally defected fingers was deployed to investigate the individual internal fingers, which suggested that the zinc-fingers recognized DNA motifs in groups: fingers 4–7 targeted the CCCTC core motif, and the flanking fingers formed the other two groups recognizing the downstream and upstream motifs[12]. While the study was carried out *ex vivo*, the DNA recognition pattern was built-up in the presence of CTCF cofactors.

Two recent studies revealed the structures of CTCF zinc-fingers in complex with different DNA fragments. One fragment was derived from the sequence at a TAD border, and the structure suggested a continuous binding of fingers 3–7 specifically to the core motif in the canonical mode. Fingers 2, 8, and 9 did not make base-specific interaction, while fingers 10 and 11 were invisible in the density map[23]. Another fragment was based on CTCF binding sites in *PCDH* locus, and the fingers 3–11 were shown in the structures. Similar DNA-recognition mode was found for the

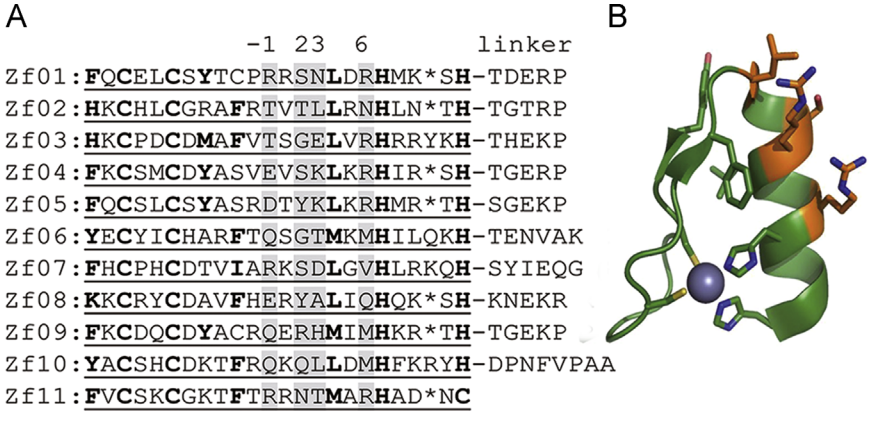


Figure 1 The sequence of the CTCF 11 zinc-fingers and the structure of a zinc-finger motif. (A) The sequence of 11 zinc-fingers. Finger motifs were underlined; cysteines and histidines for zinc-ion binding and hydrophobic residues for constructing the motif structure were in bold letters. The key residues for DNA recognition were in shed with their helical positions marked on top of them. (B) A model of zinc-finger structure motif. The residues at helical positions −1, 2, 3 and 6 were shown in orange (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

fingers 3–7, but different from the other study, the fingers 9–11 also made base-specific binding[24]. CTCF therefore recognized the two DNA targets with different patterns of effective fingers. Thus far, only the crystal structures provided the clear information of individual finger usage on a given DNA target. While the crystal structure of a multi-finger DNA complex was never a simple task, a more convenient method would accelerate the research process on multi-zinc-finger DNA recognitions.

To gain better understanding of the CTCF DNA recognition and to detect the DNA unbound regions of CTCF that would have the potential to contact its cofactors, we sought to study the role of different parts of CTCF in DNA association, particularly the individual zinc-fingers. The residues at helical positions −1, 2, 3 and 6 of zinc-finger were mutated into alanines, unless they were glycine or alanine in the wild-type protein. In this way, the finger module was inactivated in specific DNA interaction, while the structure remained intact. Only one finger was perturbed in each mutant protein, and the binding affinities were examined with various DNA probes. The affinity reduction comparing to the wild-type protein therefore reflected the effectiveness of individual fingers in DNA binding.

2. Materials and methods

2.1. Construction of plasmids for protein expression

For the cloning of full-length CTCF, total RNA was extracted from Hela cell using TRIzol reagent (Invitrogen, ThermoFisher Scientific, Carlsbad, CA) and following the manufacturer's protocol, cDNA library was constructed by reverse transcription utilizing the kit of TransScript First-Strand cDNA Synthesis SuperMix (Transgen Biotech), and CTCF-coding gene was amplified from the library by PCR with the primers listed in the Supplementary Information Table S1. The CTCF cDNA was subsequently cloned into the vector of pEASY-T-Simlpe, and verified by sequencing.

The full-length CTCF was divided into 3 regions: N-terminal (residues 1–265), zinc-finger (residues 261–582), and C-terminal (residues 578–727) regions, and the coding DNA fragments were amplified accordingly with the primers listed in the Table S1. N- and C-terminal regions were cloned into the vector of pET30a, while the zinc-finger region was cloned into the vector of pMALc2x with a TEV protease cleavage site inserted preceding the CTCF sequence. The constructs were sequenced to guarantee the correctness.

Zinc-finger mutants were achieved by site-direct mutagenesis in three steps: mutation on helix position –1, mutation on helix positions 2 and 3, and mutation on helix position 6. Primers for individual finger mutations were listed in Supplementary Information Table S2. The plasmids were mutated according to the following pattern: mZF1 (R277A, S279A, N280A, R283A), mZF2 (T305A, T307A, L308A, N311A), mZF3 (T333A, E336A, R339A), mZF4 (E362A, S364A, K365A, R368A), mZF5 (D390A, Y392A, K393A, R396A), mZF6 (Q418A, T421A, M424A), mZF7 (R448A, S450A, D451 A, V454 A), mZF8 (E478A, Y480A, Q484A), mZF9 (Q506A, R508A, H509A, M512A), mZF10 (Q534A, Q536A, L537A, M540A), mZF11 (R566A, N568A, T569A, R572A). The correctness of mutations was confirmed by DNA sequencing.

2.2. Protein expressions and purifications

Plasmid encoding either N- or C-terminal protein was transformed into *Escherichia coli* Transetta (DE3) cells, which were incubated at 37 °C in LB media (Kanamycin+, 30 mg/mL) to $OD_{600}=0.6$. Protein expression was induced by adding IPTG to reach the final concentration of 0.5 mmol/L. The cells were harvested after further growth at 20 °C overnight.

Cell pellet from 1 L culture was resuspended in 10 mL cell lysis buffer (25 mmol/L Tris pH7.5, 300 mmol/L NaCl, 5% glycerol, 1 μg/mL pepstatin A, 2 μg/mL leupeptin, 2 μg/mL aprotinin, 0.5 mmol/L PMSF), followed with sonication at 10,000 Hz for 2 min with 2 s pulse at 4 °C. Supernatant was collected by centrifugation at the speed of 40,000 × *g* for 30 min at 4 °C.

The protein-containing supernatant was loaded onto a 5 mL HiTrap chelating column, and washed thoroughly with the cell lysis buffer. The N-terminal protein was eluted with additional 100 mmol/L imidazole pH 8.0 in the buffer. The eluted protein solution was diluted to reduce the salt concentration to 200 mmol/L, before it was loaded onto a heparin column for further purification. The heparin column was washed thoroughly with the washing buffer (20 mmol/L Tris pH 7.5, 150 mmol/L NaCl, 5% glycerol, 2 mmol/L EDTA, 3 mmol/L DTT, 1 μg/mL pepstatin A, 2 μg/mL leupeptin, 2 μg/mL aprotinin, 0.5 mmol/L PMSF), before a salt concentration gradient between 150 mmol/L and 2 mol/L was applied to the heparin column. The protein was found eluted at 350 mmol/L of NaCl concentration. A final step of purification with size-exclusion column was applied to achieve a high purity, and a single sharp elution peak was observed using the running buffer identical to the heparin washing buffer. The purity was examined by SDS-PAGE, and the concentration was measured according to the UV absorbance of the protein solution.

For the purification of the C-terminal protein, only HiTrap chelating column and size exclusion column were applied with the same buffers and washing protocols. The purity was examined by SDS-PAGE, and the concentration was measured in BCA method.

The expression of CTCF zinc-finger region or its mutants was similar to that for N- or C-terminal regions, except ampicillin was supplied to the growth media instead of kanamycin. Cell lysis was achieved through the same sonication process with a slightly different cell lysis buffer (20 mmol/L Tris pH 7.5, 150 mmol/L NaCl, 5% glycerol, 1 μg/mL pepstatin A, 2 μg/mL leupeptin, 2 μg/mL aprotinin, 0.5 mmol/L PMSF, 5 mmol/L DTT, and 25 μmol/L ZnSO4).

For purification, cell lysate supernatant was loaded onto a 5 mL heparin column, and washed with a salt concentration gradient between 150 mmol/L and 1.5 mol/L. The protein was eluted at the salt concentration of 0.9 mol/L. The eluted protein was digested with TEV protease at 4 °C over 6 h, and was purified with a heparin column again to remove the MBP tag. A final step of size exclusion column was applied to further purify the protein, and a single sharp peak was achieved. The purity was examined by SDS-PAGE, and the concentration was measured according to the UV absorbance of the protein solution.

2.3. Gel retardation assays and data analysis

Single strand DNA oligos were chemically synthesized, and for each pair of oligos, one of them was labeled with the fluorophore FAM at 5′-end (Ruibio Biotech Co., Ltd.). Each oligo was dissolved in the buffer of 20 mmol/L Tris pH 8.0, 150 mmol/L NaCl, 5% glycerol to reach the final concentration of 100 mol/L. The solutions of paired oligos were mixed in equal volume, and double strand DNA fragments were generated through an annealing process: the solution of paired oligo mixture was heated at 95 °C for 15 min and let cool slowly to room temperature.

Each binding reaction was performed with 400 nmol/L protein and 40 nmol/L DNA probe in the buffer of 20 mmol/L Tris pH 8.0, 150 mmol/L NaCl, 5% glycerol, 5 mmol/L DTT, and 25 μmol/L $ZnSO_4$. Reactions were incubated at 4 °C for 1 h, before was examined on a 10 cm × 10 cm × 1 mm 7% acrylamide/0.24% bis acrylamide non-denaturing gel. The gel was run in 1 × TG (25 mmol/L Tris, 250 mmol/L glycine, pH 8.8) buffer at 10 V/cm for 80 min. The particular concentrations of protein and DNA were chosen to let 80%–90% of DNA bound with the wild-type protein, which allowed clearer observations of affinity reductions with the mutant proteins.

Gel slices were subjected to the exposure at the exciting light wavelength of 590 nm for 150 ms to capture the images with fluorescent signals (CLiNX science instruments). The fluorescent bands were subsequently measured (Chemi Analysis), while base-line was carefully calibrated in each electrophoresis lane.

Any shifted band was considered representing a portion of protein-bound DNA fragments, particularly those trapped at the loading wells of electrophoresis. Densities of all bands were combined to give out the total amount of DNA in the binding experiment, the sum of densities from the shifted bands represented the amount of DNA bound with protein, and the comparison of the two produced the percentage of shifted DNA. The individual shifted-DNA percentage was normalized against the percentage figure generated from the wild-type protein, before they were plotted together for comparisons.

2.4. *Fluorescence polarization assay*

Fluorescence polarization experiments were performed in 384-well black assay plates (Corning 3676) at a final volume of 20 μL. Briefly, 100 nm of FAM-labeled DNA fragment was incubated with a serial of 2-fold-diluted proteins in the binding buffer (25 mmol/L Tris–HCl, 50 mmol/L NaCl, 1 mmol/L DTT, 50 μmol/L $ZnSO_4$, 25 ng/μL polydI:dC and 5% glycerol) at room temperature for 30 min, and then the measurement was carried out on an ENVISON microplate reader (PerkinElmer) with polarized filters ($\lambda_{ex} = 480 \pm 25$ nm, $\lambda_{em} = 535 \pm 25$ nm). The results mP values were calculated by nonlinear regression fitting using the GraphPad Prism5.0 software.

3. Results

3.1. *Both N- and C-termini had the ability to interact with DNA, but the zinc-finger region of CTCF contributed most in DNA binding*

Based on the sequence feature of CTCF, the full-length protein was divided into 3 regions: zinc-finger region and its flanking N- and C-terminal regions. Since both N- and C-terminal regions harbored basic-residue rich sequences, which potentially had the ability to interact with negatively charged DNA molecules, we examined the differences of the DNA-binding affinities among these 3 regions.

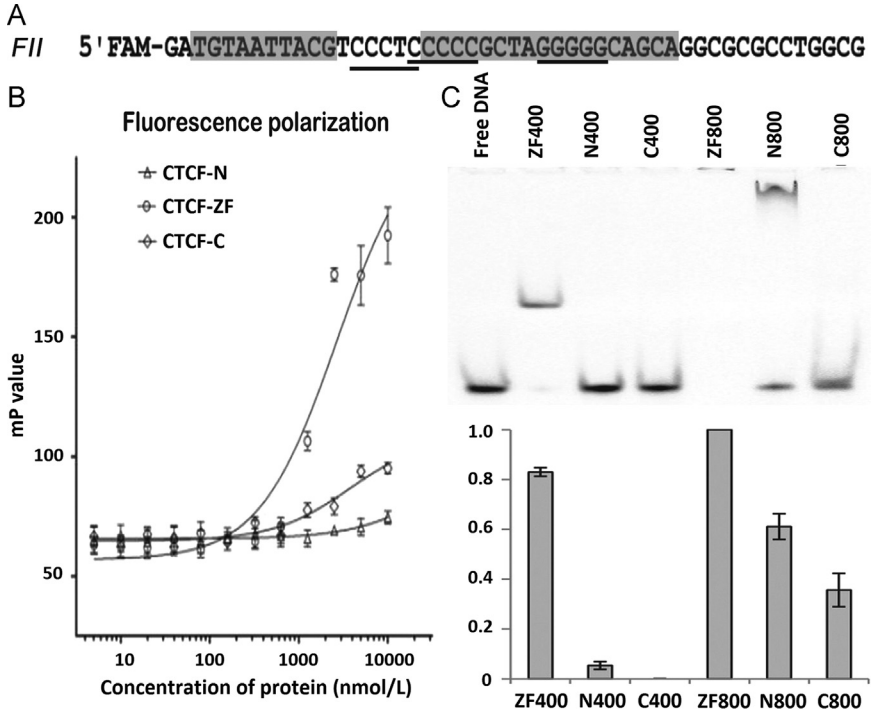


Figure 2 Affinity differences of the 3 CTCF regions. (A) Sequence of *FII* DNA probe. Three potential core motifs were underlined, and the shaded sequence had the best match with both the core and upstream motifs. (B) Fluorescent polarization experiment with the 3 CTCF region proteins. (C) Gel retardation experiment with the 3 proteins. ZF400, N400 and C400 stood for the zinc-finger, N- and C-terminal region proteins in the concentration of 400 nmol/L, while ZF800, N800 and C800 stood for the proteins in the concentration of 800 nmol/L. Standard deviations were calculated from 6 measurements.

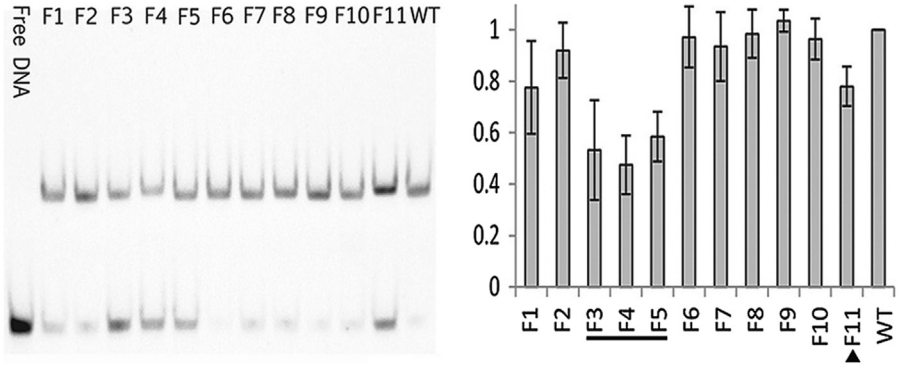


Figure 3 Affinity differences of CTCF zinc-finger mutants. Binding affinities were normalized against that of wild-type protein. Standard deviations were calculated from 11 measurements.

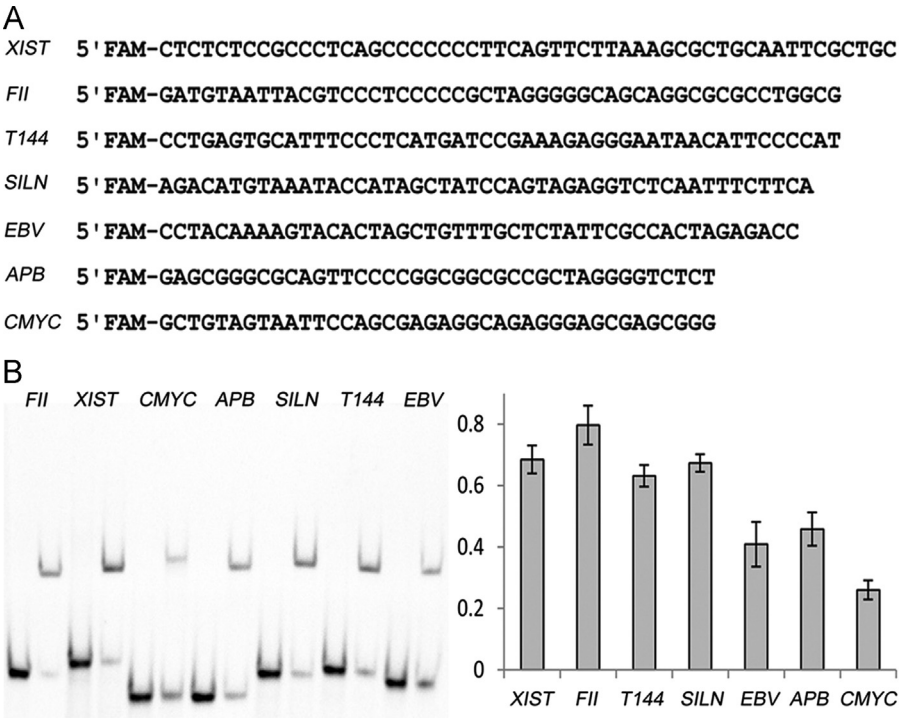


Figure 4 Sequences of the 7 DNA fragments and the comparison of their binding affinities to CTCF zinc-fingers. (A) The sequences of the 7 probes. (B) Binding affinities of 7 DNA fragments were measured under the identical condition. The vertical bars represented the ratio of shifted DNA to the total DNA, and the standard deviations were calculated from 4 measurements.

The DNA fragment for the protein binding was chosen to cover the second of the 4 CTCF-footprint regions on the chicken β-globin insulator (*FII*) (Fig. 2A), which had been extensively studied[25] and had been applied in the studies of CTCF interactions from different species[26].

With the fluorescent labeled *FII* probe, we tested the polarization signals caused by the associations of proteins. The result showed that the zinc-finger region had much higher affinity to the DNA (Fig. 2B). Additionally, we performed a gel retardation assay with the same DNA probe and proteins, but the poly-dIdC was not supplied to allow nonspecific interactions. The result was nevertheless in consistence with the fluorescent polarization experiment. The zinc-finger region caused an over 80% DNA shifting at the concentration of 400 nmol/L, while N- or C-terminal regions caused less DNA shifting even at the concentration of 800 nmol/L (Fig. 2C). Thus the zinc-finger region was the major DNA binding region of CTCF, although both N- and C-terminal regions of CTCF did have the ability to interact with the DNA target.

3.2. *Mutations on individual fingers affected DNA binding affinity to different degrees*

To investigate the contribution of individual zinc-fingers in binding *FII*, we inactivated fingers by mutating key-residues for DNA recognition while keeping the finger structure integrity intact. Based on the extensive structural and molecular biology studies, it was well established that a zinc-finger motif made DNA sequence specific recognition by inserting the N-terminal residues of its α-helix into the major groove of DNA double helix[17]. The configuration constrained the recognition most likely to happen through the contacts with the key residues at helical positions -1,

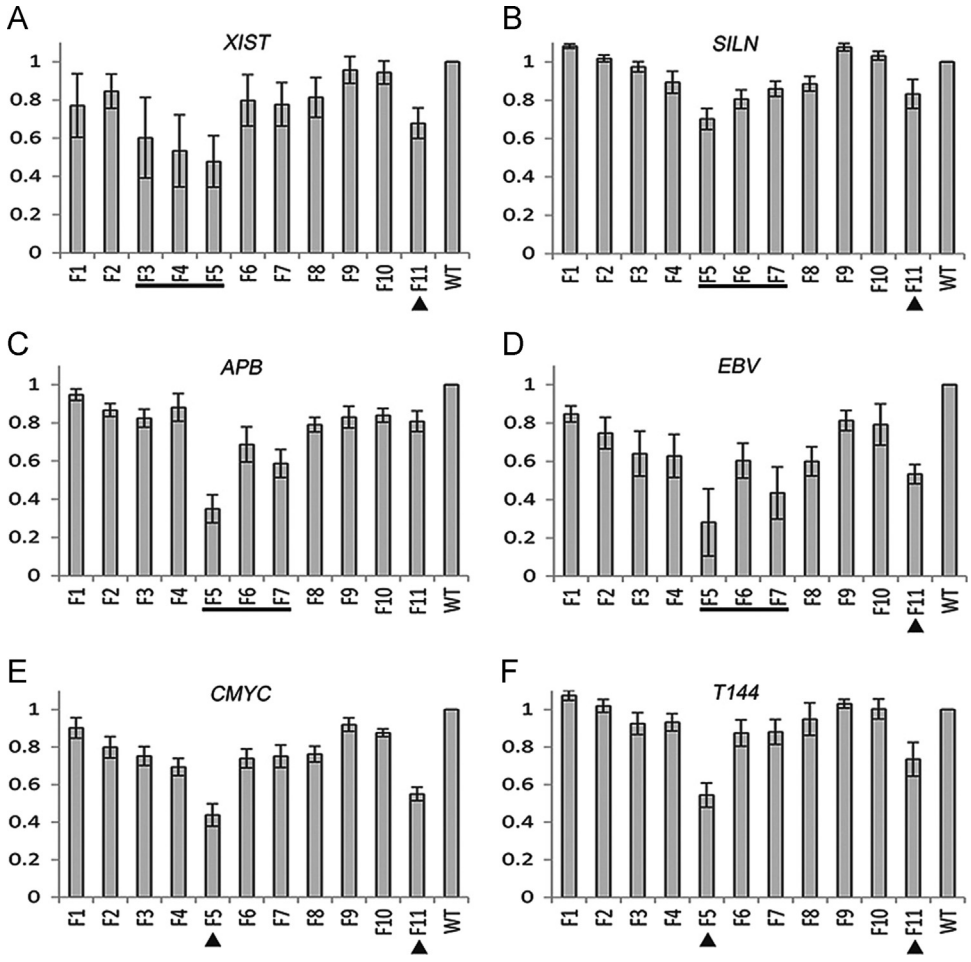


Figure 5 The interaction patterns of 11 CTCF zinc-fingers varied on 6 DNA targets. Effects of single finger mutants to the binding affinities of various DNA probes: (A) *XIST*, (B) *SILN*, (C) *APB*, (D) *EBV*, (E) *CMYC*, and (F) *T144*. DNA binding affinities were normalized against that of wild-type protein, and standard deviations were calculated from at least 4 measurements.

2, 3, and 6, and the residues at other positions are less likely to make base-specific contacts due to their spatial orientations. We mutated the 4 key residues of individual zinc-finger motifs into alanines, which were inert in DNA interactions. The mutagenesis strategy was further supported by the crystal structures of CTCF DNA complexes, which showed that these residues only conducted base-specific interactions with their DNA targets[23,24].

The 11 inactive-finger mutants were subsequently examined for their affinity alterations, taking the wild-type protein as the reference. Apart from the finger 9 mutant, which displayed a slightly higher affinity, all others showed reduced binding affinities. When the errors were considered, mutations on fingers 3–5, 1 and 11 led to lower affinities than the wild-type protein, and the other fingers had no significant effect upon inactivation (Fig. 3). Thus, the CTCF zinc-fingers recognized the *FII* probe in a discontinuous mode.

Although the pattern of effective fingers 1, 3–5, and 11 roughly followed the idea of recognition into 3-DNA-motifs, the sequence of *FII* was unable to provide indication to the recognition pattern. In the sequence, there were 3 CCC(T/C)C-alike motifs (underlined in Fig. 2A). To reach the best fit of consensus sequences would require the GGGGG present in the core motif (shaded in Fig. 2A). However, these 5 bases were not protected in the footprinting assay[25], which suggested that the core motif located more likely at the 2 other potential sites. Thus the DNA sequence-orientated estimation of CTCF zinc-finger pattern could not apply to this *FII* probe.

3.3. Sequence variation alone in the DNA targets could affect the binding affinity of CTCF zinc-fingers, but the finger recognition patterns were not predictable based on the DNA sequences

Over the vast number of DNA binding sites, different CTCF occupancies have been noticed throughout the human genome[12,27]. It was however not known whether the difference came directly from the CTCF–DNA interaction or was a consequence of the collaborations of different CTCF cofactors. We took *FII* and additional 6 DNA fragments from different sources for CTCF affinity test. They included fragments from human *XIST* gene promoter (*XIST*)[22], the promoter-proximal region of human *CMYC* gene (*CMYC*)[28], *APBβ* region in the promoter of human amyloid *β*-protein precursor (*APB*)[29], chicken lysozyme silencer (*SILN*)[20], human thyroid hormone response element 144 (*T144*)[21], and a fragment from Epstein-Barr virus (*EBV*)[30], all of which were derived from the specific CTCF protection regions in footprinting assays.

Under the same concentration of CTCF zinc-finger protein, there was indeed a clear difference in the binding affinities among the 7 DNA probes (Fig. 4). *FII* showed the highest affinity with nearly 80% of total DNA shifted, while *CMYC* had the lowest with the shifted DNA accounted for less than 30%. *APB* and *EBV* showed relatively weak binding with about 40% shifted DNA, while *XIST*, *SILN* and *T144* bound to the zinc-fingers in the high affinities with shifted DNA around 65%. Although in general shorter DNA fragments showed lower affinity, clear differences existed between those in same lengths, such as *CMYC* and *APB*, and *T144* and *FII*. *SILN* was 1 bp longer than *EBV*, but was significantly stronger in binding of CTCF zinc-fingers.

Although the affinity difference suggested possible variation in CTCF finger usage, the target DNA sequences could not provide good estimation on the pattern of effective fingers. Among the 7 DNA probes, both *XIST* and *FII* harboured multi-CCCTC motifs in their sequences, which made it difficult to predict the recognition pattern of CTCF zinc-fingers. A more puzzling situation was seen on *EBV* where no clear motif signature could be detected. To verify the exact fingers deployed for the recognition of a given DNA target, an *in vitro* experimental approach would be more efficient to address the question.

3.4. *Mutagenesis study revealed the patterns of CTCF finger effectiveness that varied on different DNA probes*

CTCF binding sites were widely spread in human genome[31]. Although the binding pattern of the CTCF 11-fingers had been classified into 8 formats based on the pattern variation of the 3 DNA motifs in the CTCF binding sites[12], for a given particular DNA site such as *FII* and *EBV*, the DNA sequence based estimation was impractical, and the experimentally verified effective/ineffective fingers would be a more direct solution. To this end, we applied the inactive-finger mutants to the above 6 DNA fragments, to examine the feasibility of the method in the identification of pattern variations of effective fingers.

When the inactive finger mutants were applied to the above 6 CTCF binding DNA fragments, different patterns were observed. Similar to *FII*, the strongest affinity reduction at *XIST* was seen on the mutations of fingers 3–5 and 11. Fingers 1, 2 and 6–8 showed milder effects, but the inactivation of fingers 9 or 10 did no harm to the *XIST* interaction (Fig. 5A). Quite different from the *FII* and *XIST*, chicken lysozyme silencer sequence (*SILN*) showed the affinity reduction with the mutations on fingers 4–8 and 11 (Fig. 5B). For *APB*, *EBV* or *CMYC*, mutations on any finger showed reduced affinity in the DNA-binding. Among them, *APB* and *EBV* were more sensitive to the mutations on fingers 5–7 (Fig. 5C–D), while *CMYC* was more sensitive to the mutations on fingers 5 and 11 (Fig. 5E). Interestingly, *T144* showed similar finger-pattern to the *CMYC*, where mutations on fingers 5 and 11 also caused strong affinity loss. In addition, fingers 3, 4, 6 and 7 had milder effects, but fingers 1, 2 and 8–10 did not show any effect upon inactivation (Fig. 5F).

The results based on the inactive finger mutants were largely in consistent with the previous finger-truncation studies, which had been performed on *XIST*, *SILN*, *APB*, *CMYC* and *T144*. The best agreements were shown on *APB* and *T144*, where the fingers 5–7 and fingers 5–11 had been thought essential for CTCF interaction, respectively[21,29]. A narrower range of effective zinc-fingers was shown previously on *SILN*, which only covered fingers 5–8[20]. The finger 4 was not included, but did show some effect once truncated. A wider range of fingers, on the other hand, were thought important for *CMYC* interaction, which included fingers 3–11[28]. While the effect of losing fingers 1–3 or 1–4 was less significant than the others, and the inactivation of the fingers 3 or 4 did cause affinity reduction (Fig. 5E), the results from the two approaches generally agreed with each other.

The largest apparent difference was shown in the analysis of effective fingers for *XIST* interaction. Fingers 6–8 were thought the most important by the finger-truncation method[22], while we found the effects from fingers 3–5 were more dominant. When all the effective fingers, strong or weak, were considered altogether, the results from the two approaches started to agree with each other: fingers 1–8 all contributed to the interaction of *XIST*, and any truncation on the first 5 fingers did cause profound loss of affinity[22]. The effect of finger 11 was nevertheless not detected previously on *XIST* and *SILN*, possibly owing to the relatively small contribution of a single finger to the total binding affinity achieved by the multiple effective-fingers.

Some fingers played more dominant roles in DNA binding, and the pattern of these fingers varied on different targets. Correlating to the core and the upstream DNA motifs, the dominant fingers could be divided into two groups, fingers 3–7 and finger 11. Different sets of fingers from the first group were required depending on the DNA targets: fingers 3–5 for *FII* and *XIST*, fingers 5 and 7 for *APB* and *EBV*, and the finger 5 alone for *SILN*, *CMYC* and *T144* (Fig. 6). This was not observed in the crystal structures, where fingers 3–7 made base-specific binding simultaneously. While the studies with finger-truncated proteins seldom

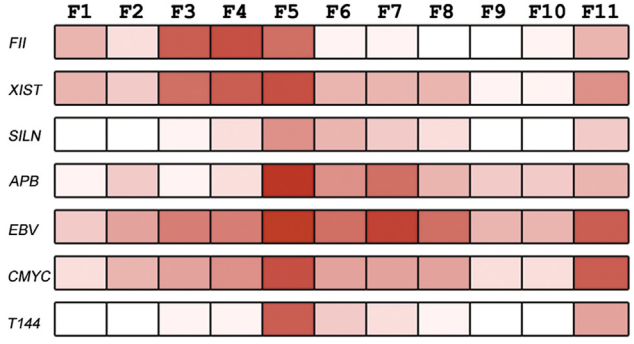


Figure 6 The heat map of interaction patterns. The degree of redness increased for every 5% affinity reduction of single finger mutants comparing to the wild-type (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

detected all these 5 fingers in an effective region, we speculated that, comparing to the dominant fingers, the other fingers in the group acted but less decisively in the DNA recognition. Another possibility was that the simultaneous binding of two DNA motifs produced a mutual constraint between the two groups of zinc-fingers, which was lost in the crystal structures characterizing the 11 fingers separately in small groups. Under the circumstance of a DNA complex with all 11 fingers, when a dominate finger was inactivated, the balance between the two motifs was broken, and CTCF suffered a drastic affinity loss to its DNA target.

4. Discussion

In this work, we examined the *in vitro* DNA-binding features of CTCF. While both its N- and C-terminal regions had weak abilities in DNA binding, it was the zinc-fingers that played the major part in the CTCF DNA association. The affinity nevertheless varied on different DNA targets, which suggested that under the identical condition, CTCF binding would bias towards some sites within a genome. The low-affinity sites may require additional conditions to gain better access to CTCF, such as higher CTCF concentration, more favorite local conformation, or the presence of other factors. We noticed that in a previous *ex vivo* study, different CTCF occupancies were found on its binding sites throughout the genome. While binding affinity could definitely affect CTCF occupancy at a given site, to what degree the effect was made by the single factor remained elusive. Interestingly, the CTCF binding sites with low-occupancy were frequently found to be cell-type specific, which was thought a result in adapting the rapid change of gene regulations during cell differentiation[27]. In this study, the viral sourced DNA probe *EBV* also displayed a weak affinity. Since the manipulation of CTCF expression could regulate the downstream genes of this DNA site[30], possibly the virus deployed a similar regulation strategy for the genes of the invaded virus in response to the host cell environment change.

By deploying inactivated-finger mutants, we examined the role of individual fingers on different DNA targets. While our results were consistent to the previous studies in general, the new method allowed us to identify ineffective fingers and dominant fingers within the previously found effective group of consecutive fingers. Furthermore, the variation was shown on the patterns of dominant effective fingers among the examined DNA targets, which suggested different binding modes. Comparing to the DNA sequence based estimation of CTCF binding mode, the dominant finger pattern avoided the potential complications of the heavy DNA sequence variations and the alike context sequences of a CTCF binding site, and provided a more direct representation to the surface feature of a DNA bound CTCF. The distinctive surface feature would subsequently permit the specific association of various CTCF cofactors and/or drug molecules, and eventually deliver the biological functions.

More than 70,000 distinct CTCF binding sites have been identified over different cell lines, and some 50,000 sites can be found in a single cell line[31]. The vast number of DNA sites need to correlate to the different biological functions of CTCF to make them meaningful. It has been shown that CTCF delivers its function in association with many other factors, such as COHESIN, OCT4, TAF3, Wrap53 and Jpx[8–10,32,33]. Some of these factors make direct contacts with DNA-bound CTCF, and some of them compete with DNA elements for the CTCF association. In both cases, the exposed surface feature of DNA-bound CTCF is important to decide which factor has the priority to interact with the CTCF at a given DNA site, and the pattern of dominant fingers can provide an indication to the surface feature.

C2H2 zinc-finger proteins formed the largest family which accounted for 3% genes of human genome[17]. Multi-finger proteins widely existed but were not well studied partly due to the limited method available to tackle the problem. Footprinting assay was the most reliable way to find the protected DNA fragment, and serial truncation mutagenesis from either N- or C-terminus of a zinc-finger chain was often deployed to characterize individual fingers. However, the truncation method was only capable to detect the fingers at the boundary of an effective region. Further deletion could not generate useful information, while one of the dominant effective fingers had already been removed. The effort was made in two approaches to investigate internal fingers. One was to delete individual fingers directly from a chain of fingers[29], and the other was to mutate zinc-ion binding residues leading to the collapse of the structure module[12]. Both methods would end up to the spacing alteration between the neighboring zinc-finger modules at the mutation site, and would potentially cause unexpected side-effects to the protein.

Here we described a mutagenesis strategy to inactivate individual zinc-fingers with intact overall structural configuration. The mutant proteins with only one inactivated zinc-finger motif subsequently permitted us to examine all zinc-finger motifs. Comparing to the finger-truncation method, the finger-inactivation method was likely to be more sensitive to detect the effectiveness of individual fingers in DNA binding, while the accumulating effect from the removal of multiple fingers did not exist. Comparing to the *ex vivo* study method, the *in vitro* method did not have the potential complications from the endogenous CTCF and its cofactors, and the experiment was better controlled quantitatively. Through the analysis with 7 DNA targets, the method was proved feasible, and the finger effectiveness in DNA binding was revealed.

With the identification of ineffective fingers and dominant fingers within a set of consecutive fingers, the strategy for further studies could be rationally designed to focus on certain region of the full-length protein, which would consequently simplify the study of CTCF interactions with its cofactors. While there were many more multi-finger proteins, some with more than 30 zinc-finger motifs, it would be interest to see whether the finger-inactivation method could be extended to the studies of other proteins.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31770804); CAMS Initiative for Innovative Medicine, China (Grant No. 2016-I2M-1–009); and the IMM Basic Research Fund, China (Grant No. 2014ZD03).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.apsb.2018.08.007.

References

1. Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 2005;**3**:e157.
2. Leiberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range

interactions reveals folding principles of the human genome. *Science* 2009;**326**:289–93.
3. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;**159**:1–16.
4. Filippova GN, Qi CF, Ulmer JE, Moore JM, Ward MD, Ju YJ, et al. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Res* 2002;**62**:48–52.
5. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 2000;**405**:482–5.
6. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* 2017;**169**:930–44.
7. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 2011;**479**:74–9.
8. Zlatanova J, Caiafa P. CTCF and its protein partners: divide and rule?. *J Cell Sci* 2009;**460**:128–32.
9. Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT. The pluripotency factor OCT4 interacts with CTCF and also controls X-chromosome pairing and counting. *Nature* 2009;**460**:128–32.
10. Liu Z, Scannell DR, Eisen MB, Tjian R. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* 2011;**146**:720–31.
11. Lobanenkov V, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaia AV, et al. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CTCF-motif in the 5′-flanking sequence of chicken *C-MYC* gene. *Oncogene* 1990;**5**:1743–53.
12. Nakahashi H, Kwon KRK, Resch W, Vian L, Dose M, Stavreva D, et al. A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep* 2013;**3**:1678–89.
13. Lu D, Searles MA, Klug A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* 2003;**426**:96–100.
14. Fan Y, Lu D. The Ikaros family of zinc-finger proteins. *Acta Pharm Sin B* 2016;**6**:513–21.
15. Choo Y, Klug A. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A* 1994;**91**:11168–72.
16. Isalan M, Klug A, Choo Y. Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* 1998;**37**:12026–33.
17. Klug A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* 2010;**43**:1–21.
18. Moore M, Klug A, Choo Y. Improved DNA binding specificity from polyzinc finger peptide by using strings of two-finger units. *Proc Natl Acad Sci U S A* 2001;**98**:1437–41.
19. Nolte RT, Conlin RM, Harrison SC, Brown RS. Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA. *Proc Natl Acad Sci U S A* 1998;**95**:2938–43.
20. Burcin M, Arnold R, Lutz M, Kaiser B, Runge D, Lottspeich F, 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**:1281–8.
21. Awad TA, Bigler J, Ulmer JE, Hu YJ, Moore JM, Lutz M, et al. Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence. *J Nutr Biochem* 1999;**274**:27092–8.
22. Pugacheva EM, Tiwari VK, Abdullaev Z, Vostrov AA, Flanagan PT, Quitschke WW, 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**:953–65.
23. Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, Cheng X. Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol Cell* 2017;**66**:711–20.
24. Yin M, Wang J, Wang M, Li X, Zhang M, Wu Q. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res* 2017;**27**:1365–77.
25. Chung JH, Bell AC, Felsenfeld G. Characterization of the chicken *β*-globin insulator. *Proc Natl Acad Sci U S A* 1997;**94**:575–80.
26. Chung JH, Whiteley M, Felsenfeld G. A 5′ element of the chicken *β*-globin domain serves as an insulator in human erythroid cells and protects against position effect in drosophila. *Cell* 1993;**74**:505–14.
27. Essien K, Vigneau S, Apreleva S, Singh LN, Bartolomei MS, Hannenhalli S. CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biol* 2009;**10**:R131.
28. Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, 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**:2802–13.
29. Quitschke WW, Taheny MJ, Fochtmann LJ, Vostrov AA. Differential effect of zinc finger deletions on the binding of CTCF to the promoter of the amyloid precursor protein gene. *Nucleic Acids Res* 2000;**28**:3370–8.
30. Chau CM, Zhang XY, Mcmahon SB, Lieberman PM. Regulation of Epstein-Barr virus latency type by the chromatin boundary factor CTCF. *J Virol* 2006;**80**:5723–32.
31. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, et al. Widespread plasticity in CTF occupancy linked to DNA methylation. *Genome Res* 2012;**22**:1680–8.
32. Saldana-Meyer R, Gonzalez-Buendia E, Guerrero G, Narendra V, Bonasio R, Recillas-Targa F, et al. CTCF regulates the human P53 gene through direct interaction with its natural antisense transcript, *WRAP53*. *Gene Dev* 2014;**28**:723–34.
33. Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT. *Jpx* RNA activates *Xist* by evicting CTCF. *Cell* 2013;**153**:1537–51.