Nucleic Acids Research, 2024, **52**, 6490–6506
https://doi.org/10.1093/nar/gkae391
Advance access publication date: 14 May 2024
Molecular Biology

Single-molecule imaging reveals a direct role of CTCF's zinc fingers in SA interaction and cluster-dependent RNA recruitment

Jonas Huber, Nicoleta-Loredana Tanasie, Sarah Zernia and Johannes Stigler*

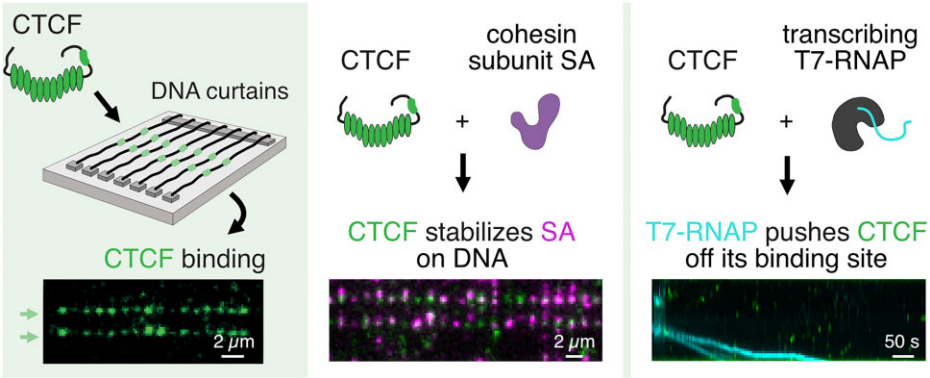
Gene Center Munich, Ludwig-Maximilians-Universität München, Munich, Germany

*To whom correspondence should be addressed. Tel: +49 89 2180 76951; Email: stigler@genzentrum.lmu.de

Abstract

CTCF is a zinc finger protein associated with transcription regulation that also acts as a barrier factor for topologically associated domains (TADs) generated by cohesin via loop extrusion. These processes require different properties of CTCF-DNA interaction, and it is still unclear how CTCF's structural features may modulate its diverse roles. Here, we employ single-molecule imaging to study both full-length CTCF and truncation mutants. We show that CTCF enriches at CTCF binding sites (CBSs), displaying a longer lifetime than observed previously. We demonstrate that the zinc finger domains mediate CTCF clustering and that clustering enables RNA recruitment, possibly creating a scaffold for interaction with RNA-binding proteins like cohesin's subunit SA. We further reveal a direct recruitment and an increase of SA residence time by CTCF bound at CBSs, suggesting that CTCF-SA interactions are crucial for cohesin stability on chromatin at TAD borders. Furthermore, we establish a single-molecule T7 transcription assay and show that although a transcribing polymerase can remove CTCF from CBSs, transcription is impaired. Our study shows that context-dependent nucleic acid binding determines the multifaceted CTCF roles in genome organization and transcription regulation.

Graphical abstract



Introduction

Eukaryotic genome architecture, as revealed by high resolution Hi-C maps (1,2), mirrors the plethora of interactions between different genomic regions. Indeed, two levels of organization have emerged: a global one represented by compartmental domains which are determined by the segregation of transcriptionally active and inactive chromatin regions within the nucleus (3–5) and a local one where loci of the same domain form strong interactions, leading to the formation of topologically associated domains (TADs) (6,7). Human CCCTC-binding factor (CTCF) is a transcription factor (8) that can be found at TAD boundaries (9). Hi-C experiments showed that removal of CTCF leads to a strong reduction in insulation between TAD domains (10), indicating that CTCF blocks loop-extruding cohesin complexes (11,12), and is hence a regulator of chromatin looping.

CTCF consists of 11 zinc fingers (ZFs), that recognize a specific DNA binding sequence (CTCF-binding site, CBS), and two unstructured termini (13,14). CBSs are diverse (15) and share a core sequence that is recognized by the central ZFs of CTCF while upstream motifs are bound by C-terminal ZFs (Figure 1A) (14,16). CBSs can be further classified into high-and low-affinity binding sites, as certain CBSs have been shown to be more persistent in CTCF depletion experiments (17,18). Low-affinity binding sites are often located within genes and at transcription start sites (TSSs), while high-affinity binding sites can be found at TAD boundaries (19). This implies that the sequence context modulates CTCF's activity on

Received: January 3, 2024. **Revised:** March 21, 2024. **Editorial Decision:** April 22, 2024. **Accepted:** May 1, 2024

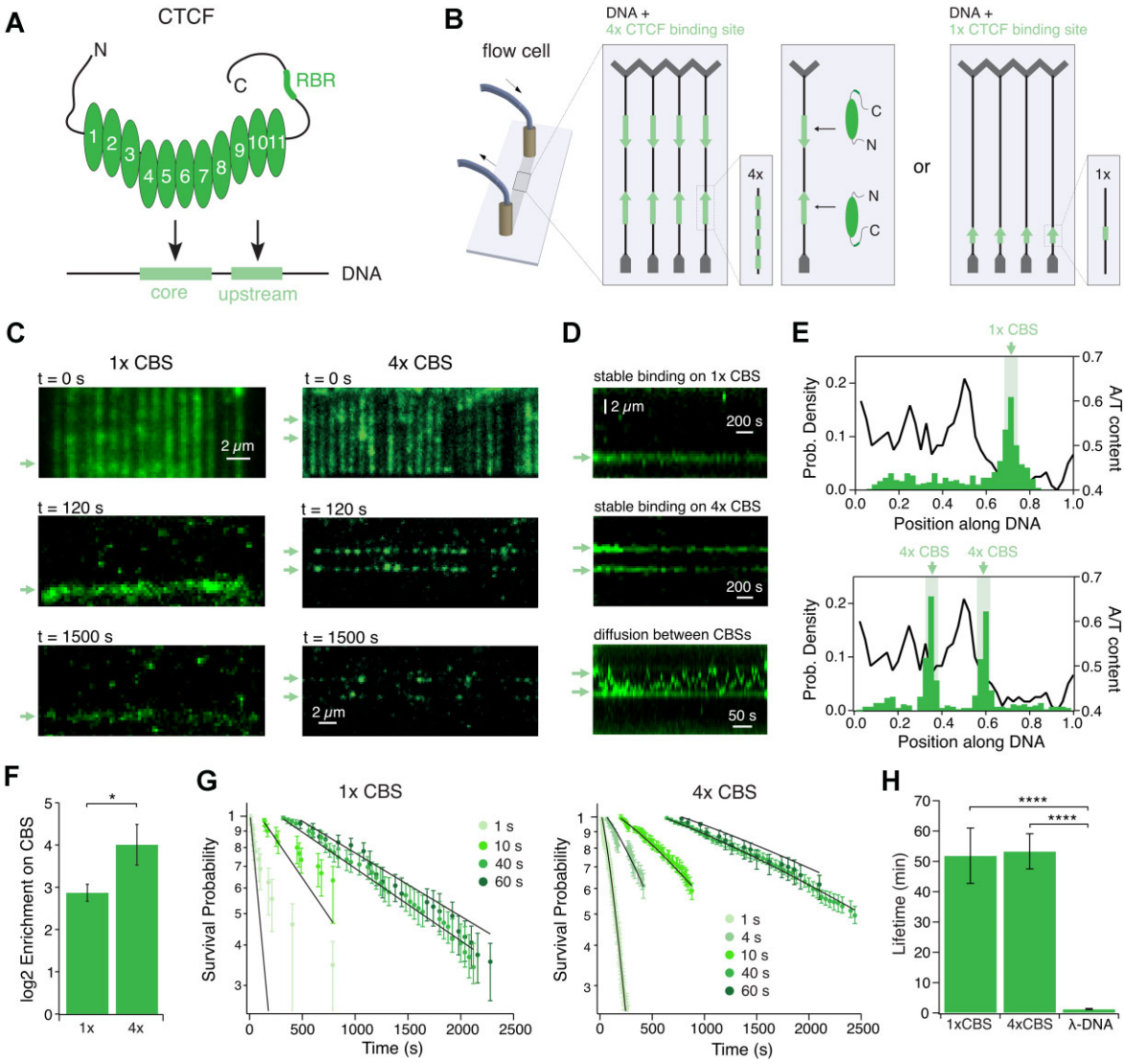


Figure 1. CTCF enriches on both, single and 4× CTCF-binding sites, with a lifetime of ∼50 min. (**A**) Schematic representation of CTCF containing the 11 ZFs, the RNA recognition motif (RBR), the elongated termini and the CTCF binding site (CBS) containing core and upstream motif. (**B**) Schematic representation of the DNA curtains assay. The DNA substrate was designed with either two cassettes of 4x CBSs (light green) with opposing orientation or one cassette with 1× CBS that were included into λ-DNA (black), which was tethered between Cr barriers (grey) on a custom-built flow cell. (**C**) Representative images of TIRF microscopy of 10 nM AF-568-CTCF-WT binding on DNA including either 1× (left) or 4× (right) CTCF binding sites (CBSs, indicated by green arrows). CTCF was first loaded at 50 mM NaCl leading to full coverage of the DNA (top). After washing with 300 mM NaCl, CTCF is enriched on the binding sites (middle) and remains bound for a long time (bottom). (**D**) Representative kymogram of CTCF binding to 1× (top) or 4× (middle) CBSs. Kymogram is shown after a 300 mM NaCl wash. Some non-CBS-bound CTCF diffuse (bottom). (**E**) CTCF enriches at 1× ($N = 477$) and 4× CBSs ($N = 427$) after a high salt wash. Light green bars indicate position of CBSs. A/T-content of the DNA substrate is shown with the black line. (**F**) CTCF is more strongly enriched (16×) on a 4× than on a 1× site (7×). (**G**) Lifetimes on 1× and 4× CBSs measured at different laser frame delay and 100 ms illumination time. A global fit was applied to correct for photobleaching and the presence of multiple CTCFs on the 4× CBSs. (**H**) Photobleaching-corrected CTCF lifetimes on 1× CBS and 4× CBSs are similar, but significantly higher than on λ-DNA.

different target sites and defines its numerous tasks in genome organization and transcription regulation.

A vast majority of CBSs located at TAD borders are convergently oriented (2,20), i.e. with the CTCF N-termini pointing towards the inside of the TADs. Cohesin is stopped when it encounters the N-terminus of CTCF. This orientation-dependent arrest of loops is determined by a direct interaction of cohesin subunits SA and Rad21 with the CTCF N-terminus (21,22). However, *in-vivo* studies revealed that SAs remain associated with CTCF even in the absence of cohesin (23), suggesting a cohesin-independent interaction between CTCF and SA. Another study showed SA to be the only cohesin subunit directly interacting with CTCF, dependent on CTCF's C-terminus, in contrast to the Rad21-dependent interaction with its N-terminus (21,24). Alternatively, the association of CTCF and SA could be facilitated by their shared ability to bind RNA (25–27). The exact determinants of CTCF-SA interaction remain therefore unknown.

CTCF interacts with RNA through both its outer ZFs ZF1 and ZF10 and an RNA binding domain (RBR) within the disordered C-terminus. This interaction seems to also modulate several aspects of chromatin organization (28,29). Furthermore, RNA association is linked to CTCF multimerization and cluster formation (26), which have also been observed

in living cells (30,31). Nevertheless, it remains unclear, which CTCF features mediate cluster formation and whether clusters contribute to CTCF's roles at TADs borders and in transcription regulation.

Complementary to its roles in chromatin looping, CTCF is also associated with gene expression regulation, as it can cooperate with other transcription factors (32). For example, the TFII-I transcription factor can bind CTCF, and recruit CDK8 for transcription initiation (33). In addition, CTCF's role as a transcription regulator is further established by a direct interaction with the largest subunit of RNA Polymerase II (Pol II) (34). CTCF also acts at later stages of gene regulation as it can influence alternative splicing by promoting Poll II pausing (35), with the exact mechanism for pausing and recruitment of additional factors thus far undetermined.

CTCF has been attributed many roles in genome regulation, yet mechanistic details of how it fulfills them remain scarce. We set out to ascertain the molecular determinants for CTCF's various tasks in genome organization and functionality. We purified fluorescently labeled full-length and truncation variants of human CTCF and visualized their association with CBSs using DNA curtains. We show that while a single CTCF is sufficient for target site binding, CTCF's ZFs mediate clustering on DNA, which enables RNA recruitment. By establishing a single-molecule transcription assay, we show that CTCF is displaced from its CBS by an elongating polymerase, leading to impaired transcription. Furthermore, we uncover a previously unknown mechanism for CTCF association with cohesin, as we show that cohesin's subunit SA can be recruited by CTCF bound at CBSs, independently of CTCF termini. Our single-molecule investigation provides insight into how ZF-mediated CTCF interaction with both RNA and DNA modulates CTCF's various roles at TAD boundaries and in transcription regulation.

Materials and methods

CTCF expression, purification and labeling

Long CTCF constructs (CTCF WT, ΔN, ΔC, ΔNC, ΔRBR) were expressed in High Five (Hi5) insect cells. To this regard we cloned cassettes containing the respective CTCF variant with an N-terminal 6xHis- and Halo-tag and a C-terminal Flag-tag into a pLIB expression vector (kindly provided by Karl-Peter Hopfner, LMU Munich). Expression vectors were transformed into DH10MultiBac cells to create a bacmid. Virus amplification was repeated two times in Sf21 cells, before transfecting 1×10^6 High Five cells/ml with a 1:500 virus dilution. Protein expression continued for 3 days at 27°C before harvesting the cells and freezing them in liquid N_2. Cells were resuspended in CTCF resuspension buffer (25 mM Hepes pH 8.3, 100 mM NaCl, 5% glycerol, 0.05% tween, 100 μM $ZnCl_2$, 10 mM imidazole, 1 mM TCEP, 1 mM PMSF, protease inhibitor tablet (Roche)) and sonicated for 90 sec. Short CTCF constructs (ZF4-7, ZF9-CT) were cloned into a pET28a vector containing an N-terminal 6xHis-tag and a C-terminal Flag-tag. The plasmid was transformed into *Escherichia coli* Rosetta (DE3) and cells were grown to an OD of 0.6 before inducing expression with 1 mM IPTG for 16 h at 18°C. Cells were harvested, resuspended in CTCF resuspension buffer and sonicated for 20 min. Lysates for long and short constructs were treated with 2000 units Pierce universal nuclease (Thermo Scientific) for 2 h at RT before increasing the salt concentration to 500 mM NaCl. Lysates were centrifuged for 1 h at 42 krpm and 4°C. Supernatants were filtered through 0.22 μm PES filters and incubated with 3 ml Ni-NTA beads (Macherey-Nagel) for 45 minutes at 4°C. The supernatant was applied to a gravity flow column and beads were washed with 100 ml of nickel CTCF wash buffer (25 mM Hepes pH 8.3, 1000 mM NaCl, 5% Glycerol, 0.05% Tween, 100 μM $ZnCl_2$, 10 mM imidazole, 1 mM TCEP) before eluting with 10 ml nickel CTCF elution buffer (25 mM Hepes pH 8.3, 500 mM NaCl, 5% glycerol, 0.05% Tween, 100 μM $ZnCl_2$, 500 mM imidazole, 1 mM TCEP). CTCF constructs were dialyzed against CTCF 500 buffer (25 mM Hepes pH 8.3, 500 mM NaCl, 5% glycerol, 0.05% Tween, 100 μM $ZnCl_2$, 1 mM DTT) before further purification on a Cytiva HiTrap Heparin HP column with a 500 to 1500 mM NaCl gradient. CTCF fractions were pooled and concentrated to 500 μl. For labeling, chloroalkane linker containing fluorescent dyes were generated by click-chemistry. 2.5 mM Halo-DBCO (Iris-Biotech) was incubated for 4 h at 37°C with 7.5 mM AF568-azide (Lumiprobe) or Atto 643-azide (ATTO-TEC GmbH). Alternatively, Halo Tag Alexa Fluor 660 (Promega) was used. CTCF was treated with 10× excess of dye for 15 min at RT followed by a final purification step on a Cytiva Superose 6 column in CTCF 500 buffer. CTCF fractions were pooled, aliquoted, flash frozen in liquid N2, stored at −80°C and thawed directly before DNA-curtains measurements.

SA1/SA2 and Rad21 expression and purification

Cohesin's subunits SA1 and SA2 were expressed in Hi5 insect cells. The cassettes containing the respective SA variant with an N-terminal 10xHis-tag and a C-terminal ybbR-tag were cloned into pLIB expression vectors. Expression and cell harvesting were performed as described for CTCF. Cells were resuspended in SA resuspension buffer (20 mM Hepes pH 7.5, 100 mM NaCl, 10% Glycerol, 15 mM imidazole 1 mM TCEP, 1 mM PMSF, protease inhibitor tablet (Roche)) and sonicated for 90 s. Lysates were treated with 2000 units Pierce universal nuclease (Thermo Scientific) for 45 min at RT before increasing the salt concentration to 300 mM NaCl. Lysates were centrifuged for 1 h at 42 krpm and 4°C. Supernatants were filtered through 0.22 μm PES filters and incubated with 3 ml Ni-NTA beads for 45 minutes at 4°C. The supernatant was applied to a gravity flow column and beads were washed with 100 ml of nickel SA wash buffer (20 mM Hepes pH 7.5, 2 M NaCl, 10% glycerol, 40 mM imidazole, 1 mM TCEP) before eluting with 10 ml nickel SA elution buffer (20 mM Hepes pH 7.5, 300 mM NaCl, 10% glycerol, 400 mM imidazole, 1 mM TCEP). SA constructs were dialyzed for 1 h at 4°C against SA 300 buffer (20 mM Hepes pH 7.5, 300 mM NaCl, 10% Glycerol, 1 mM DTT) before further purification on a Cytiva HiTrap Heparin HP column with a 300–1000 mM NaCl gradient. SA fractions were pooled and concentrated to 500 μl. For ybbr-tag labeling, Sfp (made in house, plasmid was kindly provided by the Gaub-lab, LMU Munich) was mixed with SA in a molar ratio of 1:1 and a 1.25 excess of LD655-CoA dye (Lumidyne) as well as 10 mM $MgCl_2$. Reaction was performed for 16 h at 4°C. Proteins were purified on a Cytiva Superose 6 column in SA 300 buffer. Protein fractions were pooled, frozen in liquid N_2, stored at −80°C and thawed directly before measurements. For SA-Rad21 constructs, a Rad21 peptide consisting of amino acids 281–420 was fused to a N-terminal 6xHis- and MBP-tag in a pET28a vector and expressed in *E. coli*

Rosetta (DE3) cells as described above for small CTCF constructs. Harvested cells were resuspended in SA resuspension buffer and sonicated for 20 min. Lysates were treated with 2000 units Pierce universal nuclease for 45 minutes at RT before increasing the salt concentration to 500 mM NaCl and centrifugation at 4°C and 17 krpm for 30 min. Supernatants were filtered through 0.22 μm PES filters and incubated for 45 minutes at 4°C with 3 ml Ni-NTA beads. The supernatant was applied to a gravity flow column and beads were washed with 100 ml of nickel SA wash buffer before eluting with 10 ml nickel Rad21 elution buffer (40 mM Hep pH 7.5, 500 mM NaCl, 400 mM Imidazole, 5% glycerol, 1 mM TCEP). Proteins were dialyzed at 4°C for 1 h against Rad21 500 buffer (40 mM Hep pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT) and concentrated to 500 μl. Rad21 and SA1 or SA2 (after Heparin) were then mixed in a molar ratio of 1:1 and filled up to 500 μl with SA 300 buffer. Heterodimers were labeled and purified like SA.

Generation of λ-DNA constructs for DNA curtains

Wild-type λ-DNA was purchased from NEB. For generation of 2x T7-4×CBSs λ-DNA, a cassette containing a total of four 129 bp spaced CTCF binding sites (2 × 5′-TGCAGTTCCAAAACTGGCCAGCAGAGGGC ACCAAA-3′ and 2 × 5′-TGCAGTTCCAAAAGCGGCC AGCAGGGGGCGCCCAA-3′) and a 2700 bp upstream T7 promoter site (5′-TAATACGACTCACTATAGG-3′) was cloned at two positions in opposite orientation into λ-DNA using NgoMIV/XbaI and XhoI/NheI sites. For generation of 1×T7-1×CBS λ-DNA a single CTCF site (5′-TGCAGTACCAACTTTAACCAGCAGAGGGCACCAAA-3′) was cloned into the XhoI/NheI site and a single T7 promoter into the NgoMIV/XbaI site. The products were then packaged into phage particles using phage extract (MaxPlax, Epicentre) and amplified by lytic growth in LE392 cells (NEB). Following lytic growth, λ-DNA was purified by PEG-precipitation and phenol-chloroform extraction before resuspension in TE 150 buffer (10 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl). DNA ends were tagged with either biotin (5′-aggtcgccgccc-bio-3′) or digoxigenin (5′-gggcggcgacct-dig-3′) containing oligonucleotides (Metabion) by hybdridization and ligation to cos sites and purified on a HiPrep 16/60 Sephacryl S-300 HR column in TE 150 buffer.

Single-molecule DNA curtains experiments

DNA curtain experiments were carried out as described previously (36) on a prism-type TIRF microscope (Nikon Eclipse Ti2), equipped with three illumination lasers (488, 561 and 640 nm Coherent OBIS), an electron multiplying charged coupled camera (iXon Life, Andor) and a syringe-pump-driven microfluidics system supplying the sample chamber. Custom made flow cells were assembled from silica-fused slides grafted with chromium barriers produced via E-beam lithography and cover slips with a double-sided tape.

Lipid Master Mix preparation was adapted from a previous protocol (36). In brief, 100 mg DOPC (Avanti 850375P-200mg) dissolved in 1 ml chloroform, 1 ml DOPE-PEG (Avanti 880130C-25mg) and 50 μl DOPE-biotin (Avanti 870273C-25mg) were mixed and stored at −20°C. 100 μl Master Mix was dried using N_2 followed by applying a vacuum for 1–2 hours and subsequently resolved in 2 ml lipid buffer (10 mM Tris pH 7.5, 200 mM NaCl, 20 mM $MgCl_2$). Lipids were sonicated 5 times for 1 min with 1 min pauses on ice (amplitude 20%, duty cycle 20%), filtered through 0.22 μm PVDF filters and stored at 4°C for 3–4 weeks. Lipids were diluted 1:10 in lipid buffer and flow cells were incubated three times for 10 min with 200 μl diluted lipids. After washout, flow cells were incubated with 2 μl 1 mg/ml anti-digoxigenin (produced in house) in 700 μl lipid buffer followed by 5 μl streptavidin (Carl Roth) in 1 ml BSA buffer (40 mM Tris pH 7.5, 1 mg/ml BSA, 1 mM $MgCl_2$). After this, 0.3 pM λ-DNA in BSA buffer was added in four steps with 5 minutes incubation time. Single-molecule measurements were performed in CTCF 50 buffer (40 mM Tris pH 7.5, 1 mg/ml BSA, 1 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT) including an oxygen scavenger system containing glucose-oxidase (Carl Roth), catalase (Sigma Aldrich) and 0.4% glucose. Videos were recorded in NIS Elements (Nikon) and analyzed in Igor Pro 8 (Wavemetrics) using custom written code.

CTCF on DNA curtains

10 nM CTCF was incubated for 30 s on the DNA curtains to cover the DNA substrate, followed by CTCF enrichment on the CBS by a high salt buffer wash using CTCF 300 buffer (40 mM Tris pH 7.5, 1 mg/ml BSA, 1 mM $MgCl_2$, 300 mM NaCl, 1 mM DTT) for 3 min. For this, CTCF 300 buffer was flushed in, and flow was then stopped for incubation time. For lifetime measurements, this incubation was increased to up to 60 minutes. ZF9-Cterm was loaded at 50 nM and ZF4-7 at 200 nM concentration. Both were labeled with Flag-QD705.

CTCF was imaged using a 561 nm laser at 50 mW (0.35 $\mu W/\mu m^2$) for lifetime and diffusion videos and 140 mW (0.98 $\mu W/\mu m^2$) for photobleaching analysis. For lifetime analysis, 100 ms illumination time and a frame delay of 1, 4, 10, 40 or 60 s was used (Supplementary Figure S1A).

Cohesin's subunits SA1 and SA2 on DNA curtains

100 nM SA1, SA2, SA1-Rad21 or SA2-Rad21 were incubated for 3 min on the flow cell in CTCF 50 buffer or CTCF 150 buffer and videos were recorded with 100 ms illumination time and 2 ms frame delay using a 640 nm laser at 140 mW (0.98 $\mu W/\mu m^2$).

SA and CTCF on DNA curtains

For sequential load experiments, first CTCF and then 100 or 400 nM SA were loaded as described above and videos were recorded with 100 ms illumination time at 1s frame delay using a 561 nm laser at 50 mW (0.35 $\mu W/\mu m^2$) and a 640 nm laser at 140 mW (0.98 $\mu W/\mu m^2$). For combined loading experiments, 100 or 400 nM SA was preincubated with 10 nM CTCF in CTCF 100 buffer (40 mM Tris pH 7.5, 1 mg/ml BSA, 1 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT) for 5 min at RT before 30 s incubation in CTCF 50 buffer on the DNA curtains and 3 min enrichment in CTCF 300 buffer. For lifetime measurements, incubation in CTCF 300 buffer was increased to up to 15 min.

Transcription experiments on DNA curtains

Transcription experiments were performed by incubation of 0.33 mM ATP, CTP, GTP, UTP (NEB), 0.03 mM Cy3-UTP (Jena Bioscience), and 1:300 T7 RNA polymerase mix (NEB) in transcription buffer (30 mM Tris pH 7.5, 2 mM spermidine, 25 mM $MgCl_2$, 10 mM DTT) for 3 min at 40°C, before loading for 2 min on DNA curtains. After wash-out, a 50 μM or 1 mM ATP, CTP, GTP, UTP (missing labeled UTP) solution

in transcription buffer was loaded and a video was recorded at 100 ms illumination time and 1 s frame delay using a 561 nm laser with 25 mW (0.18 μW/μm^2) laser power.

In case of CTCF transcription experiments, CTCF-Alexa Fluor 660 was loaded and enriched as described above before starting T7-Pol transcription as described above. Transcription was recorded using 1 s frame delay and 25 mW (0.18 μW/μm^2) of 561 nm laser to detect produced RNA and 10 s frame delay and 70 mW (0.49 μW/μm^2) of 640 nm laser to visualize CTCF at 100 ms illumination time. For SA transcription experiments, SA-LD655 was incubated with unlabeled CTCF and enriched on CBSs as described above. Then, the two-step transcription was started. Transcription was recorded using 1 s frame delay and 25 mW (0.18 μW/μm^2) of 561 nm laser to detect RNA and 1 s frame delay and 140 mW (0.98 μW/μm^2) of 640 nm laser to visualize SA at 100 ms illumination time.

RNA interaction on DNA curtains

For RNA recruitment experiments, 100 bp Cy3-UTP labeled RNA was generated using a PCR-product containing a T7-promoter site and the HiScribe T7 High Yield RNA Synthesis Kit (NEB) as well as Cy3-UTP (Jena Bioscience). CTCF was incubated on DNA curtains with or without high salt wash and 25 ng/μl RNA in CTCF 50 buffer was added to the curtains and incubated for 3 minutes. CTCF was imaged using the 640 nm laser line at 70 mW (0.49 μW/μm^2) and 10 s frame delay, RNA using the 561 nm laser line at 50 mW (0.35 μW/μm^2) and 1 s frame delay at 100 ms illumination time. Photobleaching experiments were carried out with both laser lines at 140 mW (0.98 μW/μm^2) and 100 ms illumination time with 2 ms frame delay. RNA recruitment to SA was performed accordingly. First, SA was incubated with DNA curtains as described above. Second, 25 ng/μl RNA in CTCF 50 buffer was added to the curtains and incubated for 3 minutes. SA1 was imaged using the 640 nm laser at 140 mW (0.98 μW/μm^2) and 1 s frame delay and 100 ms illumination time.

For RNA transcript interaction experiments, transcription was performed on DNA curtains as described above. Afterwards T7-Pol was washed off for 3 minutes in CTCF 1000 buffer (40 mM Tris pH 7.5, 1 mg/ml BSA, 1 mM $MgCl_2$, 1000 mM NaCl, 1 mM DTT). SA or CTCF were then loaded as described above. Lifetime measurements were performed in CTCF 150 buffer for SAs and CTCF 300 buffer for CTCF using the 640 nm laser line at 70 mW (0.49 μW/μm^2), the 561 nm laser line at 25 mW (0.18 μW/μm^2), 1 s frame delay and 100 ms illumination time.

Data analysis for DNA curtains assay

Enrichment, survival and photobleaching data

DNA curtains data was analyzed in Igor Pro 8 using custom written code. Binding positions were determined in relation to positions of chromium anchors and barriers and divided into 40 bins. Enrichment on CTCF binding sites was calculated by determining the amount of CTCF or SA molecules in the bins containing CTCF binding sites in comparison to all other bins. Photobleaching data was analyzed according to a published method (37).

SA survival data was calculated from the fraction of SA-molecules staying bound to DNA during a 3-minute wash.

Lifetimes

To determine lifetime data unbiased by photobleaching, the time of disappearance of fluorescent spots was recorded and turned into a survival curve using the Kaplan–Meier method (38). As individual CTCF molecules could not be spatially separated on 4× CBSs constructs, a model was developed that describes the disappearance of the last molecule on a multi-binding-site array. The disappearance can be caused either by photobleaching or by dissociation from the binding site. The parameter λ is hence given by

$$\lambda = (\tau^{-1} + r^{-1}\tau_v^{-1})^{-1},$$

where r is the dimensionless ratio between off-times of the laser illumination and the observation times, and the parameter τ_v is the lifetime of the fluorescent dye until photobleaching at 100% illumination. With this, the cumulative probability density function until the disappearance of the i-th fluorophore from a fluorescent spot is

$$C_i(t) = \left(1 - \exp\left(-\frac{t}{\lambda}\right)\right)^i$$

The fit model to the observed survival curve is then

$$S(t) = 1 - \sum_{i=1}^{N} \frac{C_i(t) - C_i(c)}{1 - C_i(c)} \frac{B(i, l, N)}{1 - B(0, l, N)}$$

where c is the cut-off time after which the observation was started, l is the pre-determined labeling ratio and N is the number of binding sites ($N = 4$ or $N = 1$ in our case). The binomial distributions $B(\ldots)$ in the second term of the sum account for incomplete labeling. This fit model was verified using Monte Carlo simulations.

The experimental data was then fitted to this model with τ_v shared between all data sets, and r as common parameters for all measurements at the same frame delay.

The experimental data was then fitted to this model with the photobleaching rate τ_v as a common parameter; the laser-off ratio r was set as a common parameter for all measurements at the same frame delay, the number of binding sites N and labeling ratio l as a common parameter for measurements with the same DNA, the true CTCF lifetime τ for measurements with the same construct and the shortest reliably measured lifetime c individually for each measurement. The lifetime of RNA and SA molecules were fitted to a single exponential model.

Movement analysis

For analysis of molecular movement, individual fluorescent spots were localized and tracked as described (39,40). Molecules were split into two fractions (CBS-bound versus not CBS-bound for CTCF and bound at ≥50% AT versus <50% AT for SA). Diffusion coefficients were determined according to a published method (41).

Transcription speeds were determined from the onset of transcription until stalling/fall off or bleaching for each individual polymerase. Early stopping events were defined by transcription termination within the first 10 kb; snapback events were defined as a sudden upward movement by at least 0.5 kb. Polymerases that did not stop permanently were counted as continuously transcribing.

Bulk transcription assay

DNA constructs for bulk transcription assays were generated by PCR from a pET28a vector or from λ-DNA containing 4× CTCF binding sites followed by subsequent PCR purification. DNA constructs contained a T7-promoter followed by either 100 bp DNA (no CTCF sites), 5093 bp DNA (4× CBSs upstream motif facing T7 first after 2672 bp) or 1039 bp DNA (4× CBSs opposite orientation first after 196 bp). Reactions using equimolar amount of DNA were carried out in presence or absence of 50 nM CTCF using a HiScribe T7 High Yield RNA Synthesis Kit (NEB), adding 100 μM Cy3-UTP. Following DNAse treatment, products were purified using an RNA purification kit (NEB), absorption at 260 nm was determined and length of the RNA product was analyzed on a 1% agarose gel.

Mass photometry

Mass photometry measurements were performed on a Refeyn OneMP mass photometer with an image size of 10.8 μm × 2.9 μm. Cover slips (Roth) were cleaned by sonication in isopropanol and dried with nitrogen. Silicon culture well gaskets (Merck) were placed on the slide and 19 μl buffer was pipetted into the well. After focusing the laser, 1 μl protein was added to the buffer droplet to a final concentration of 50 nM and mixed by pipetting. Measurement was performed for 1 minute using Acquire software (Refeyn). For data analysis, Discover software (Refeyn) was used to convert measured contrasts into molecular masses (calibration was done using protein standards with known molecular mass). A histogram with 100 bins was calculated from all measured masses and Gaussian fits were applied to individual protein peaks using Igor Pro 8 (Wavemetrics).

Reagents

A list of recombinant protein sequences can be found in the supplement. For protein purification Pierce universal nuclease (Thermo Scientific #88702), Ni-NTA beads (Macherey-Nagel #74500.100), HiTrap Heparin HP column (Cytiva #10298944) and Superose 6 column (Cytiva #15377653) was used. For protein labeling Halo Tag Alexa Fluor 660 (Promega #G8471) was used. Alternatively, dyes were generated by click chemistry (AF568-azide (Lumiprobe #A5820), Atto 643-azide (ATTO-TEC GmbH #AD 643-41), 2.5 mM Halo-DBCO (Iris-Biotech #RL-3670.0025)). For ybbR-tag labeling Sfp (made in house, plasmid was kindly provided by the Gaub-lab, LMU Munich) and LD655-CoA dye (Lumidyne custom synthesis) was used. For generation of λ-DNA constructs λ-DNA (NEB #N3011S) and phage extract (MaxPlax, Epicentre #MP5110) was used. A prism-type TIRF microscope (Nikon Eclipse Ti2), equipped with three illumination lasers (488, 561 and 640 nm Coherent OBIS), an electron multiplying charged coupled camera (iXon Life, Andor) was used for single molecule experiments. Flow cells were generated with an Avanti lipid mixture (Otto Norwald), streptavidin (Carl Roth #6073.1), glucose-oxidase (Carl Roth #6028.1), catalase (Sigma Aldrich #C100-50MG) and anti-digoxigenin (produced in-house). For transcription experiments HiScribe T7 High Yield RNA Synthesis Kit (NEB #E2040S) and Cy3-UTP (Jena Bioscience #NU-821-CY3) was used. Mass photometry was performed on a Refeyn OneMP mass photometer.

Biological resources

Organisms and strains

E. coli DH10MultiBac (kindly provided by Prof. Karl-Peter Hopfner, LMU Munich); *E. coli* Rosetta (DE3) (Novagen), Spodoptera frugiperda (Sf21) insect cells (IPLB-Sf21AE) and Trichoplusia ni High Five insect cells (Invitrogen).

Plasmids

pLIB expression vector (kindly provided by Prof. Karl-Peter Hopfner, LMU Munich); pET28a (+) vector (Novagen); pLENTI CMV TRE3G Puro-CTCF WT (kindly provided by Prof. Christof Gebhardt, Ulm University); pCK-Sfp (https://www.addgene.org/159617/ kindly provided by the Gaub-lab, LMU Munich) and pAC4-ybbR-His-Dig10.3-AviTag (kindly provided by the Gaub-lab, LMU Munich and originally created by the Baker-Lab, University of Washington).

Statistical analysis

For lifetime data and SA diffusion data a two-tailed z-test was used. Counted single-molecule events for passing/blocking/binding SA or CTCF molecules as well as processive/impaired transcription events were analyzed using fisher's exact test. All other data was analyzed using a two-tailed *t*-test. A table containing number of molecules and *P*-values for all relevant experiments can be found in the supplement (Supplementary Table S1).

Results

CTCF has a significantly higher lifetime on CBSs compared to unspecific DNA

CTCF's target site consists of a core motif bound by ZFs 4–7 and an upstream motif bound by ZFs 9–11 (14) (Figure 1A). CTCF displays increased lifetime on its binding site and off-target diffusion (11,16). However, it is controversial whether a single CBS is sufficient for CTCF target-site recognition (11,16). Multiple CBSs create robustness for blocking loop extrusion (42) and highly conserved topologically associated domains (TADs) in mice contain arrays of CBSs (43), suggesting that closely spaced CTCF molecules may strengthen TAD borders, perhaps by facilitating protein-protein interactions. Interestingly, CTCF clustering has been shown to occur both *in vivo* (30,31,44) and *in vitro* (26,44). To test for CTCF binding cooperativity at the single molecule level, we generated two different λ-DNA constructs (Figure 1B): One construct containing two 4× CBSs (individual binding sites spaced by 129 bp) in opposite orientation, spaced by 13 kbp resembling a TAD, and the second construct containing only one CBS. End-modified λ-DNA molecules were attached to a lipid-bilayer and imaged using TIRF-microscopy. Fluorescently labeled CTCF was bound to DNA under low-salt conditions, which led to a complete coating of the DNA by the protein (Figure 1C, top). Unspecifically bound CTCF molecules were washed off by 300 mM NaCl, while CBS-bound CTCF remained for more than 25 minutes (Figure 1C, middle + bottom). Target-site-bound CTCF molecules mostly remained static on the CTCF binding sites even at these high salt concentrations, while single remaining non-CBS-bound CTCF molecules started to rapidly diffuse on DNA (Figure 1D, Supplementary Figure S1B). Photobleaching experiments revealed that neither diffusive nor static CTCF forms higher

oligomeric structures (Supplementary Figure S1C). After the 300 mM salt wash, a clear enrichment of CTCF on 1× and 4× CBSs was observed (Figure 1E), and enrichment was significantly higher for 4× CBSs (Figure 1F). We wondered if this is due to a higher lifetime on 4× CBSs, which would suggest a cooperative binding mechanism. To exactly determine CTCF's lifetime on its target site and correct for photobleaching we performed measurements at different frame delay for both the 1× and the 4× CBSs (Supplementary Figure S1A, Figure 1G). The data was then analyzed with a global fit model (see methods). For our analysis we assumed three free parameters: the photobleaching lifetime of the fluorescent dye, the lifetime of CTCF on DNA and the chance of each individual site on the 4× CBSs to be occupied with labeled CTCF (see Methods). The global fit resulted in no significant lifetime difference between 1× and 4× sites (Figure 1H). CTCF therefore does not bind cooperatively to multiple closely spaced binding sites, even though the distance between sites here (129 bp) is much shorter than the median distance between TAD-associated CTCF sites *in vivo* (5.3–5.9 kb in mice) (43). The lifetime was more than 40x higher on both 1× and 4× CBSs compared to λ-DNA.

CTCF's inner and outer ZFs but not its termini are required for CBS recognition

We next wanted to identify the roles of the unstructured CTCF termini and the ZFs in CTCF target site recognition. To this end, we produced different truncation mutants missing one (ΔN, ΔC) or both CTCF termini (ΔNC) or some of the ZFs (ΔRBR, ZF9-CT, ZF4-7, Figure 2A, see supplement for sequences). All mutants containing all 11 ZFs (ΔN, ΔC, ΔNC) were enriched to a similar extent on the CTCF binding sites as the CTCF WT (Figure 2A and B). In contrast, mutants containing only the upstream-motif-binding ZFs (ZF9-CT), or the core-motif-binding ZFs (ZF4–7) were not enriched on the CTCF binding site. Instead, CTCF's outer ZFs showed a clear preference for AT-rich regions on λ-DNA leading to a negative enrichment on the GC-rich CBSs (Figure 2A and 2B). Interestingly, a mutant we termed ΔRBR, which lacks the outer ZFs (ZF1, ZF10 and ZF11), that might take part in RNA binding, and the C-terminal RNA binding domain (RBR) (26,28,29), was enriched on the CBSs to a similar extent as CTCF WT (Figure 2B).

To test if the different truncation mutations influence CTCF's stability on DNA, we performed lifetime measurements at different frame delay, similarly to the WT measurements (Figures 2C and S2A). Mutants missing one or both of CTCF's termini but containing all 11 ZF showed no significant difference compared to CTCF WT (Figure 2D), while ZF9-CT had a 25-fold and ZF4-7 a 90-fold reduced lifetime. Interestingly, ΔRBR also showed a 2.5-fold reduced lifetime, although it was enriched similarly to WT. These data therefore suggest, that the ZFs which recognize the core or upstream motif (16,45,46) are not stable on the CBS on their own. Instead, binding needs to be stabilized by a combination of core-motif-binding- and outer ZFs. ZFs 9-11, known to bind to an upstream motif (45) with a higher AT-content (50%) compared to the CTCF-core motif (28%), thereby preferentially engage AT-rich regions. Minor and major groove dimensions of AT-rich DNA (47,48) might be more favorable for interactions with positively charged residues in non-core motif binding zinc fingers (16). The presence of AT-rich regions in close proximity to the CBS has been linked to long persistence times of CTCF on chromatin (18). CTCF missing only some (ZF1, ZF10, ZF11) but not all of these outer ZFs can still recognize its CBS but with a reduced lifetime. Similarly, while ZFs 5-7 were shown to be essential for binding, the deletion of ZF1, ZF10 + 11 or ZF9 leads to a significant CTCF half-life reduction on a CBS (49). In contrast to CTCF's outer ZFs, CTCF-termini including the RBR on the C-terminus are not required to increase binding stability of the core-motif-binding ZFs.

Cohesin's SA subunits specifically recognize CBS-bound CTCF in absence of other cohesin subunits

After showing enrichment of CTCF on both 1× CBS and 4× CBSs, we asked if CTCF can, according to its role in TADs formation (9), recruit different parts of cohesin to the CTCF binding site. Most recent data suggest that CTCF blocks loop extrusion via interaction of a conserved domain in CTCF's N-terminus with cohesin subunits SA and Rad21 (21). However, CTCF has also been shown *in vivo* to colocalize with SA in the absence of Rad21 (23). The exact mechanism of CTCF's interaction with SA remains unknown. We sought to test if SA1 or SA2 interact directly with CTCF on the CBSs in the absence of other cohesin subunits. To this end, we purified and fluorescently labeled SA1 and SA2. We then preincubated them with different CTCF constructs before performing a salt wash to remove off-target CTCF (Figure 3A). SA1 and SA2 were enriched with CTCF on the CBSs 10-fold or 8-fold, respectively (Figure 3B and Supplementary Figure S3A), clearly showing that SA subunits can be recruited by CTCF to the CBSs in the absence of other cohesin subunits. Strikingly, even before the salt wash, SA interacted more frequently with CBS-bound CTCF (Figure 3A, top and Supplementary Figure S3F). However, this enrichment effect was lost when SAs were loaded at four times higher concentrations (Supplementary Figure S3F). SAs are therefore able to differentiate between CBS-bound and unspecifically bound CTCFs in a concentration-dependent manner, possibly due to conformation changes in the ZFs when CTCF is able to engage a complete sequence motif.

SAs target AT-rich regions on DNA

To directly observe if cohesin's subunit SA can specifically recognize CTCF once it has already bound its CBS, we first enriched CTCF on the CBSs and then loaded SA1 or SA2 (Figure 3C). Surprisingly, this led to a completely different binding behavior. Instead of colocalizing with CTCF, both SA1 and SA2 showed a clear preference for AT-rich DNA regions (AT-content ≥ 50%) (Figure 3D). A possible explanation could be that pre-bound CTCF blocks GC-rich regions. We therefore tested the binding behavior of SA without CTCF and again found a clear preference for AT-rich regions (Figure 3E, F). By comparing lifetimes (Figure 3G) and diffusion (Figure 3H, Supplementary Figure S3B) at low salt on AT-rich (AT content ≥ 50%) and GC-rich (< 50%) regions, we found this enrichment to happen by faster diffusion on GC-rich regions and by a lower lifetime on GC-rich regions leading to more rapid sampling. At higher salt concentration (150 mM NaCl), SA quickly unbound from DNA (Figure 3G). SA diffusion was also observed on DNA previously enriched with CTCF (Supplementary Figure S3C). When SA collided with a CBS-bound CTCF, we observed a large fraction of intermediate and long-term binding events (>20 s, Supplementary Figure S3D). Less frequently, SA was blocked by CTCF and changed its

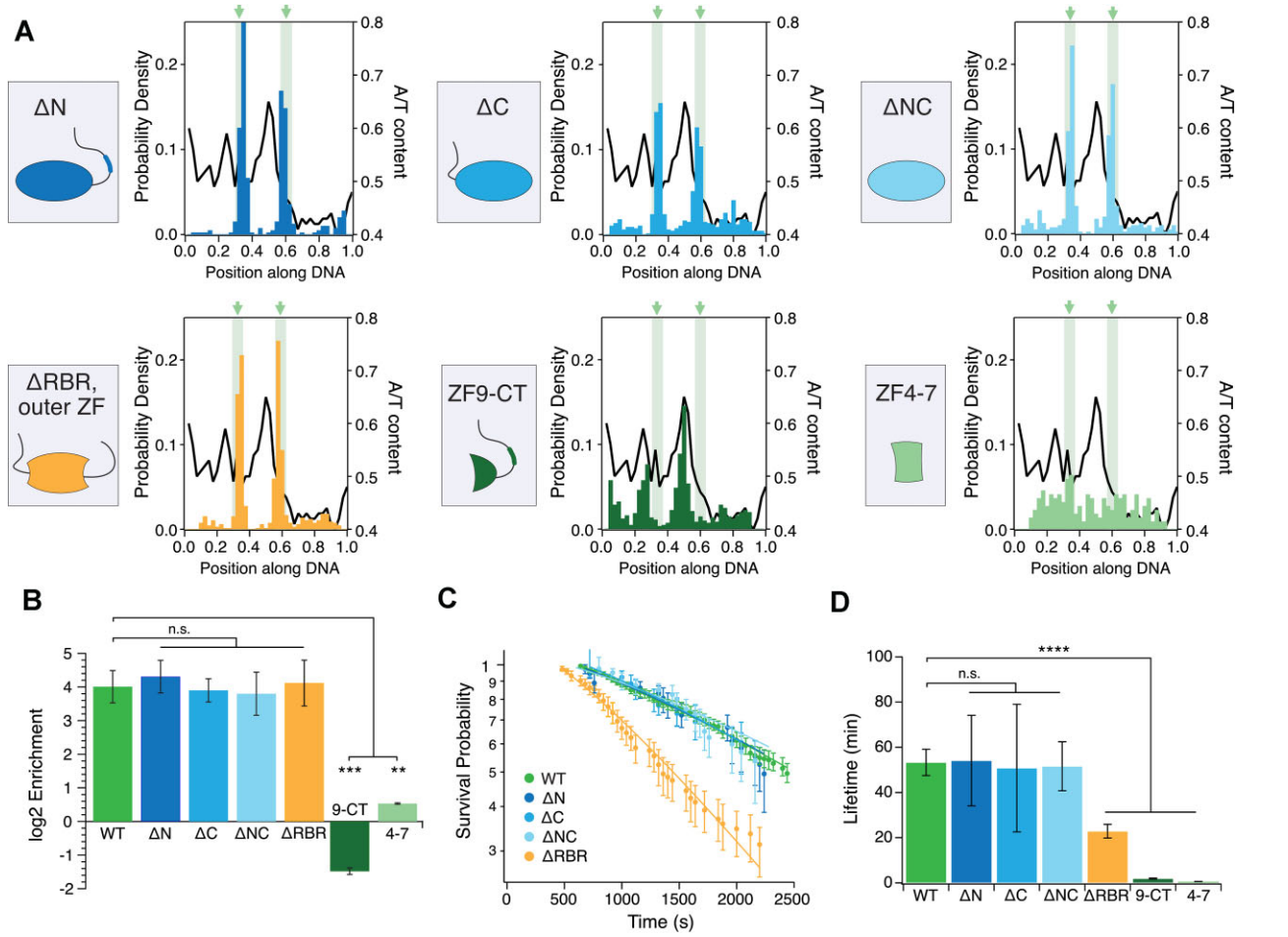


Figure 2. Binding behavior of CTCF variants on DNA curtains (**A**) Binding position of CTCF variants on 4× CTCF binding sites after 300 mM NaCl wash ($N = 390/666/396/651/343/88$ for ΔN/ΔC/ΔNC/ΔRBR/ZF9-CT/ZF4-7). ΔN/ΔC/ΔNC/ΔRBR labeled with AF568 and loaded at 10 nM concentration. ZF9-CT/ZF4-7 labeled with Flag-QD705 were loaded at 50 and 200 nM concentrations, respectively. (**B**) Enrichment of CTCF variants on 4× CBSs compared to λ-DNA. ΔN, ΔC, ΔNC and ΔRBR+ΔZF1,10,11 (ΔRBR) enrich like WT on CBSs. ZF9-CT (9-CT) and ZF4-7 (4–7) are significantly less enriched than CTCF WT. (**C**) Lifetimes of CTCF variants at 40 s frame delay and 100 ms illumination time. (**D**) Photobleaching-corrected lifetimes of CTCF variants. ΔN, ΔC and ΔNC display similar values like WT. ΔRBR, ZF9-CT and ZF4–7 display lower lifetimes.

diffusion direction. In very rare cases, we observed SA passing a CBS-bound CTCF (Supplementary Figure S3C and D). We conclude that diffusing SAs can recognize CBS-bound CTCFs on DNA. However, since our DNA substrate contains large AT-rich regions, where SA accumulates, this recruitment is comparatively rare. This high preference for AT-rich DNA regions is also observed when performing sequential load experiments at four times higher SA concentrations (Supplementary Figure S3E, F).

SAs interact more stably with CTCF's ZFs than with DNA

Since diffusion to the CBSs seems to be an inefficient mechanism of SA recruitment, we asked how cohesin's subunit SA is enriched on the CBSs in our combined loading experiments. In contrast to the low salt stability of SA on DNA, SA preincubated with CTCF and recruited to the CBSs stayed bound for multiple minutes at 300 mM NaCl (Figure 3I). This argues that CBS-associated SAs are not bound to the DNA but directly attached to CTCF.

We next sought to find out if this interaction is further stabilized by cohesin's Rad21 subunit. We therefore copurified SA1 or SA2 with a Rad21 peptide containing its CTCF binding region (21) and validated its interaction with SA by mass photometry (Supplementary Figure S3G). Rad21 did not influence SA enrichment on CTCF in sequential or simultaneous load experiments and had no influence on SA lifetimes on CTCF (Supplementary Figure S3H–K). Rad21 therefore does not seem to be required for the CTCF-SA interaction.

It is known that a conserved region in the N-terminus of CTCF is required for cohesin interaction (21,29,50). To test whether this is also true for the interaction with SA, we incubated CTCF truncation variants with SA1 or SA2 and loaded them on DNA curtains. No significant differences in interaction times (Figure 3G) or enrichment (Supplementary Figure S3A) were observed for any truncation mutant. We therefore conclude that SA is recruited to and stabilized on CBSs by specifically binding to CBS-bound CTCF via a ZF-mediated interaction.

CTCF and CTCF-SA complexes do not block transcription

CTCF has a well-established role as a transcriptional insulator, blocking distant enhancer–promoter interactions in verte-

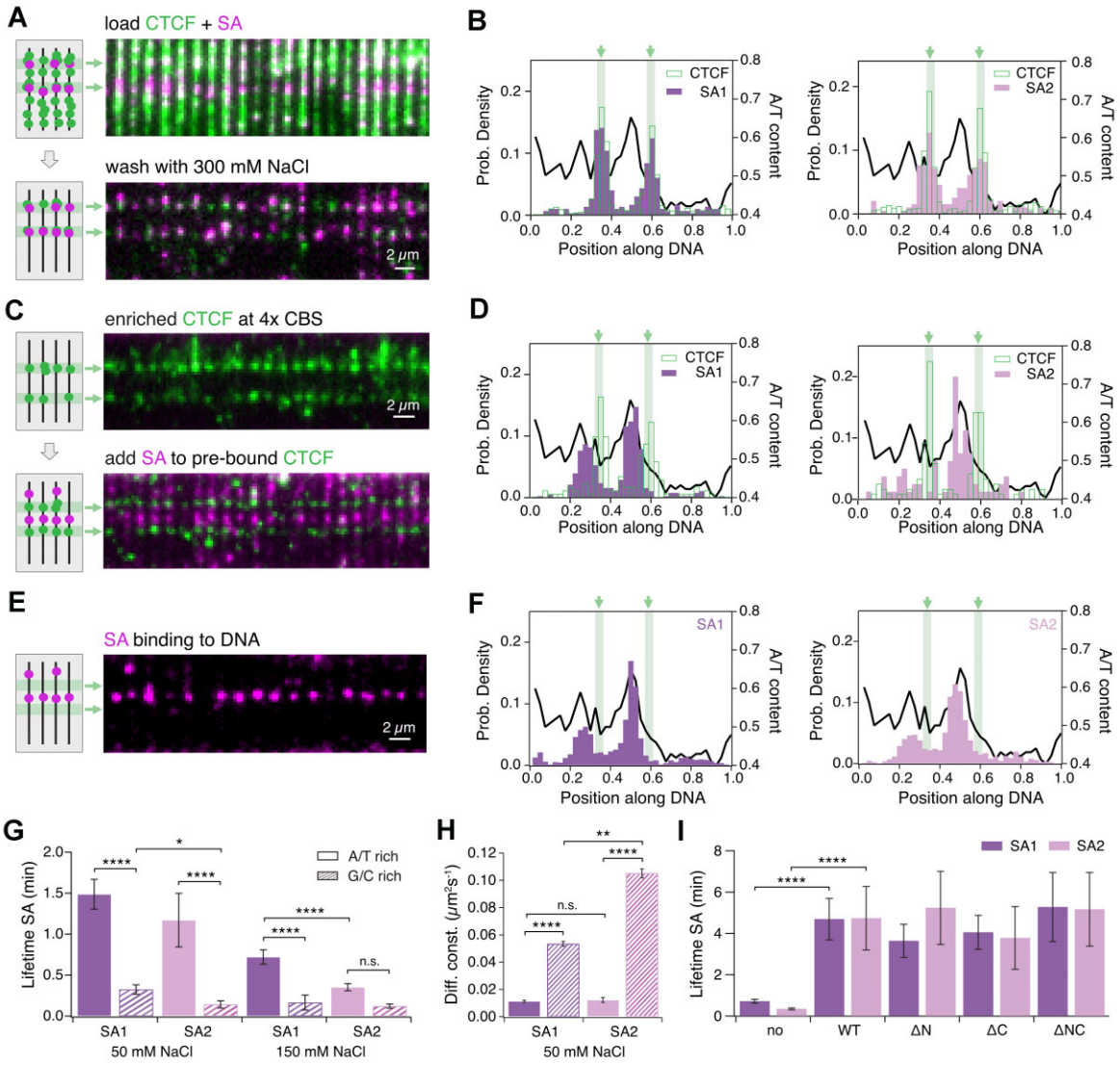


Figure 3. CTCF ZFs stabilize SA on CBSs (**A**) TIRF microscopy at 100 ms illumination time of fluorescently labeled CTCF (green) and SA (pink) preincubated (10 nM AF568-CTCF with 100 nM SA-LD655) before loading on DNA curtains, before (top) and after (bottom) 300 mM NaCl wash. (**B**) Histograms of CTCF (green, left: $N = 762$; right: $N = 747$), SA1 (purple, $N = 299$) and SA2 (pink, $N = 196$) binding positions after simultaneous load and salt wash. CBSs are shown as green bars and the AT-ratio as a black line. (**C**) TIRF microscopy of 10 nM AF568-CTCF after salt wash (top) followed by 100 nM SA-LD655 load (bottom). Color code as in (A). (**D**) Histograms of CTCF (green, left: $N = 171$; right: $N = 120$), SA1 ($N = 380$) and SA2 ($N = 80$) binding positions after sequential load. (**E**) TIRF microscopy of 100 nM SA-LD655 (magenta) binding on DNA-curtains. (**F**) Histograms of SA1 ($N = 1321$) and SA2 ($N = 875$) binding positions on λ-DNA. (**G**) Lifetime of SAs on AT-rich and GC-rich DNA regions at 50 and 150 mM NaCl. At 50 mM NaCl, both SAs have a higher lifetime on AT-rich than on GC-rich regions. At 150 mM, only SA1 shows a higher lifetime on AT-rich, which is also significantly higher than for SA2. (**H**) SA1 and SA2 diffusion coefficients on AT-rich and GC-rich regions (coloring like in G). SA1 and SA2 diffuse significantly faster on GC-rich than on AT-rich regions. SA2 diffuses significantly faster than SA1 on GC-rich regions, but not on AT-rich regions. (**I**) Lifetime of SAs in absence of CTCF and after recruitment to CBSs by different CTCF constructs. SA1 and SA2 have a significantly higher lifetime on CTCF WT than on DNA. Both SAs have similar lifetimes on CTCF variants compared to CTCF WT.

brates (51,52). CTCF might also directly influence transcription by interaction with the RNA-Polymerase II via its C-terminal domain (34). Binding of CTCF near TSSs can influence Pol II processivity *in vivo*, which impacts mRNA splicing efficiency (35,53–55). Our goal was to test both the influence of CBS-bound CTCF and CTCF-SA complexes on transcription and the influence of a transcribing polymerase on these complexes. Hence, we performed *in-vitro* transcription assays both in bulk and single-molecule using T7-RNA-polymerase (T7-Pol). For bulk measurements, we used two linear constructs containing construct-centered 4× CBSs in two orientations downstream of a T7-promoter (Figure 4A, B). Unexpectedly, neither the amount nor the length of produced RNA was influenced by CTCF on both type of templates, suggesting that CTCF does not inhibit transcription (Figure 4A, B). CTCF was therefore either passed, pushed or removed from the DNA by the transcribing polymerase.

To test these possibilities, we performed single-molecule transcription assays. For this, we altered our previously used λ-DNA-construct by inserting a T7-promoter in front of each 4× CBSs (Figure 4C). Transcription assays were then performed in two steps. First, transcription was carried out for 2 min in the presence of labeled nucleotides (Cy3-UTP) to produce fluorescently labeled RNA. Second, after a brief wash to

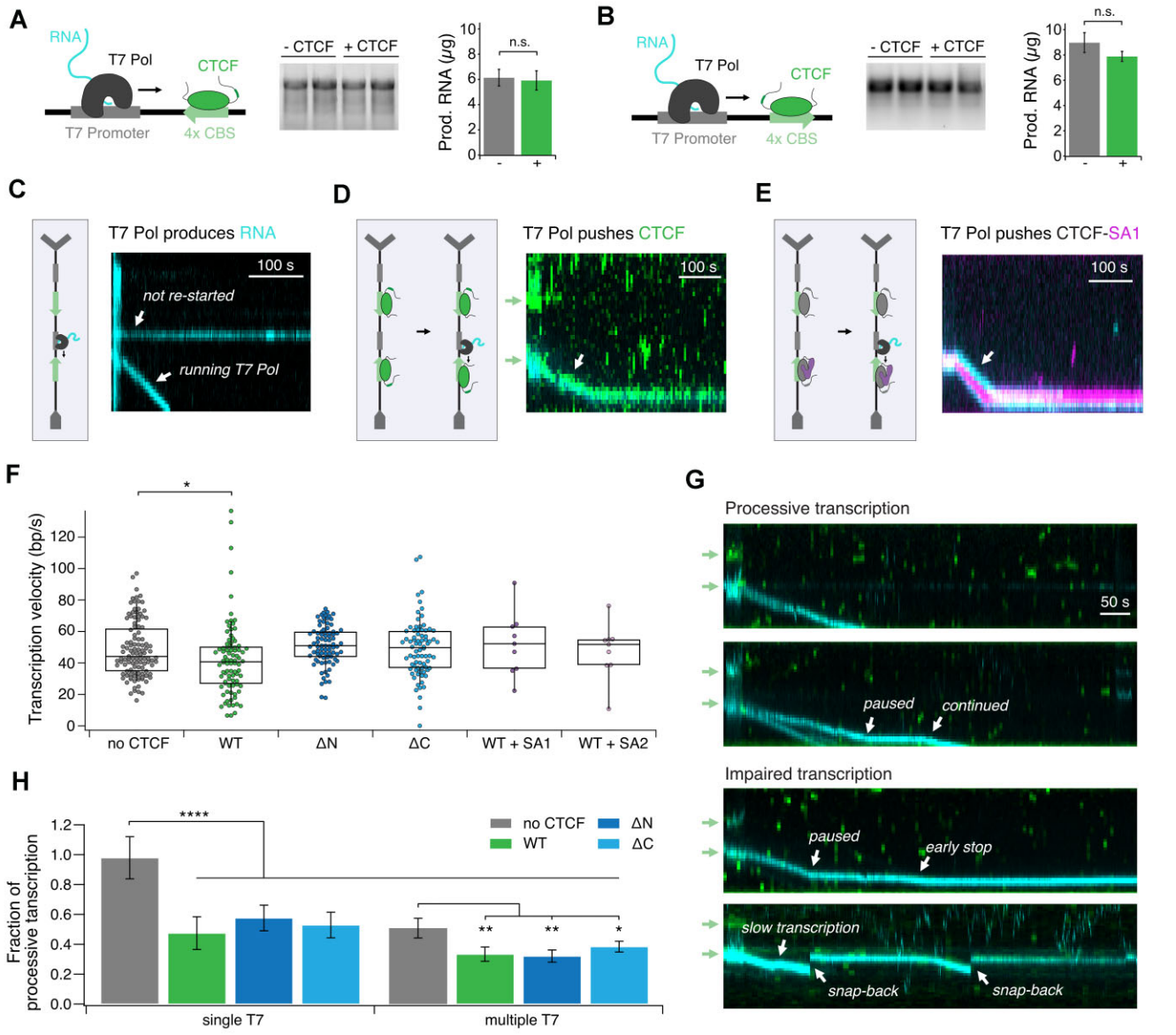


Figure 4. CTCF does not block transcription by T7 polymerase. (**A**) T7-Pol bulk transcription assay on a PCR product containing a T7-promoter as well as 4× CBSs with the downstream motif pointing towards the promoter. The length and amount of produced RNA was measured for three independent experiments with and without CTCF and no significant reduction in amount of produced RNA was observed. (**B**) Same as (A) but with upstream motif pointing towards the promoter. (**C**) Illustration of *in-vitro* transcription assays and representative kymograms for T7-Pol moving down the DNA during transcription with 1 mM nucleotide concentrations (cyan = Cy3- RNA) measured at 100 ms illumination time. (**D**) same as (C) but after enrichment of 10 nM Alexa Fluor 660-CTCF on 4× CBSs. CTCF (green) and T7-Pol (cyan, RNA) are moving mutually. (**E**) same as (C) but after enrichment of SA-CTCF complexes containing 10 nM unlabeled CTCF and 100 nM SA-LD655 (magenta) on 4× CBSs. CTCF-SA complexes are pushed by T7-Pol. (**F**) Mean transcription velocities of T7-Pol alone and T7-Pol pushing different CTCF variants or CTCF-SA complexes. WT CTCF reduced transcription velocity, while no significant difference was found for ΔN, ΔC, SA1-CTCF and SA2-CTCF compared to T7-Pol velocities. (**G**) Representative kymograms of continuous transcription (top), pausing and stopping events (middle), and snapback of polymerases (bottom). (**H**) Fraction of processive transcription in case of single or multiple transcription events on one λ-DNA molecule ($N = 49/118$ T7 control, 40/147 WT, 33/134 ΔN and 34/177 ΔC for single T7/multiple T7 transcription events). All CTCF variants impair transcription significantly in both cases of single and multiple T7 polymerases on one DNA. In absence of CTCF, multiple T7-Pols on one DNA impair transcription also significantly.

remove the labeled nucleotides, transcription was restarted in the presence of only dark nucleotides to reduce the fluorescent background in the flowcell. The movement of the transcribing T7-Pol was followed by tracking the Cy3-UTP labeled RNA (Figure 4C).

To test CTCF's influence on transcription we loaded and enriched fluorescently labeled CTCF on the CBSs as described. After restart, transcribing T7-Pols pushed CTCF off its binding site. CTCF (green) and RNA (cyan) fluorescence co-localized during the whole transcription process (Figure 4D). We compared the velocities of these pushed complexes to those of T7-Pol alone and found a small but significant reduction in transcription velocity in the presence of CTCF WT ($P = 0.034$) (Figure 4F). In contrast, the ΔN and ΔC variants did not slow down T7-Pol. Since T7 RNAP is a rapidly transcribing RNA polymerase (56) we repeated experiments at below-saturation nucleotide concentrations to be closer to RNA Pol II transcription speeds (57)

(Supplementary Figure S4A, B). T7-Pol is still able to push CTCF off its binding site (Supplementary Figure S4A) and transcription speed does not significantly change in presence of pushed CTCF complexes (Supplementary Figure S4B). We therefore conclude that CTCF does not block transcription.

Our single molecule data revealed a stable interaction between CBS-bound CTCF and cohesin's subunit SA (see above). We therefore wondered whether SA-CTCF complexes show a different behavior than CTCF alone when encountered by a transcribing polymerase. However, SA-CTCF complexes (SA labeled, CTCF dark) were also pushed by T7-Pol (Figure 4E). We conclude that neither CTCF alone nor SA1-CTCF nor SA2-CTCF complexes block transcription.

CTCF increases the frequency of impaired transcription events

In the absence of CTCF, most T7-Pol transcription elongation events (98%) were processive, with T7 polymerase transcribing more than 10 kbp of λ-DNA without permanently stopping, unbinding or restarting (Figure 4H). Strikingly, in the presence of CTCF WT, ΔN or ΔC, we observed a significant reduction of processive transcription to 48–58% and an increased amount of permanent early stopping and snap-backs (Figure 4G). After being pushed off its CBS, CTCF might get evicted from DNA by the nascent RNA chain. CTCF-RNA binding has been linked to the formation of transcription clusters *in vivo* (31). We therefore hypothesized that impaired transcription in our experiment is a consequence of increased T7-Pol collision events in CTCF-RNA clusters. To test if clustered or closely spaced polymerases impair transcription (58), we split our data into DNA molecules on which we only observed a single transcription event, versus multiple transcription events. Interestingly, even in the absence of CTCF, multiple transcribing polymerases on the same DNA led to a significant decrease in processive transcription to 51%, which is even further reduced in presence of CTCF to 32–38% (Figure 4H). We thus propose that CTCF increases interactions between polymerases by forming local clusters with RNA (26), which then disturbs transcription.

CTCF oligomerization is required for secondary RNA-capture

CTCF increases the frequency of polymerase stalling and reduces the fraction of processive polymerases (see above). We therefore asked if this processivity reduction stems from DNA-bound CTCF binding to the nascent RNA, i.e. secondary RNA capture. To test this possibility, we enriched CTCF on 4× CBSs as shown above (Figure 1C, E, F) and incubated with fluorescently tagged RNA in solution (Figure 5A). Unexpectedly, we observed no RNA recruitment to CBS-bound CTCF. CTCF-RNA binding has been linked to CTCF clustering (26,28). However, whether RNA binding is required for oligomerization or *vice versa* remains elusive. To test if CBS-bound CTCF is monomeric, we performed photobleaching experiments (Figure 5B). We found that most fluorescent puncta at 4× CBSs bleached in less than four steps, with an average of 2.5 ± 0.1 steps (Figure 5C). At a pre-determined labeling efficiency of 52 ± 8%, this is consistent with monomeric CTCF. CTCF cluster formation might require a high local concentration of DNA-bound CTCF (44). We therefore repeated the RNA capture experiments before removing unspecifically bound CTCF. In contrast to our previous result, RNA was efficiently captured and even remained bound after a high-salt wash (Figure 5D). CTCF photobleaching analysis at positions of captured RNA revealed higher oligomeric structures with an average of 8.8 ± 0.3 steps, corresponding to an average of 17 CTCF molecules (Figure 5E, F).

To analyze the role of the CTCF-termini in multimerization and RNA-capture, we repeated the same experiments for the CTCF ΔNC mutant. Interestingly, the mutant still formed multimers on DNA (9.0 ± 0.5 bleaching steps), which in turn were able to capture RNA (Supplementary Figure S5D). Hence, the termini of CTCF are not required for RNA capture and oligomerization is caused by ZF interactions (16,44). On the contrary, CTCF's termini might even impede cluster formation to some extent since the bleaching step histogram of CBS-enriched ΔNC displays a second peak not observed for any other mutant (Supplementary Figure S2B).

WT- and ΔNC-RNA clusters were enriched 4-fold on the 4× CBSs, but this enrichment was significantly smaller than for monomeric CTCF without RNA (Supplementary Figure S5A,C). More CTCF-RNA clusters stayed bound at unspecific positions, implying that RNA binding additionally stabilizes CTCF on low-affinity binding sites. In contrast to CTCF, DNA-bound SAs did not capture RNA (Supplementary Figure S5E). Instead, both cohesin subunits SA1 and SA2 were mostly washed off the DNA by RNA, indicating a higher affinity of SAs for RNA than DNA (Supplementary Figure S5F) (25). We conclude that while RNA binding removes SA, it stabilizes clustered CTCF and does not bind to monomeric CTCF.

Both CTCF and SA colocalize with RNA transcripts

Since CTCF is involved in the formation of transcriptional condensates (31), we wanted to test if CTCF's ability to bind RNA causes direct recruitment by RNA transcripts. We performed transcription experiments like described above, followed by CTCF incubation. We found that CTCF colocalized with RNA transcripts (Figure 5G, Supplementary Figure S5B), but there was no significant decrease in CBS enrichment, as CTCF prefers CBSs over RNA (Figure 5H). Lifetime on CBSs was more than 25 times larger than on RNA (Figure 5I).

In contrast, when we repeated this experiment with SAs, which colocalize with R-loops *in vivo* (23), the presence of RNA transcripts on DNA curtains led to a complete change in binding behavior (Supplementary Figure S5G, H). Both cohesin subunits SA1 and SA2 were no longer enriched on AT-rich regions (Supplementary Figure S5I) and instead accumulated at RNA transcripts, with a higher lifetime compared to DNA (Supplementary Figure S5J). This is consistent with our previous observation of SA preferring RNA over DNA (Supplementary Figure S5F). We conclude that both CTCF and SA can be recruited by RNA transcripts. However, while this recruitment drastically changes the DNA localization of SA, CTCF is still mostly localized at CBSs.

Discussion

CTCF's DNA binding dynamics are influenced by RNA and transcribing polymerases

CTCF influences a wide variety of processes in mammalian cells by binding to its genomic target site, including insulation (51,59,60), alternative splicing (35,54,61), transcription activation (62) and TAD boundary formation (6,7,9). Our study

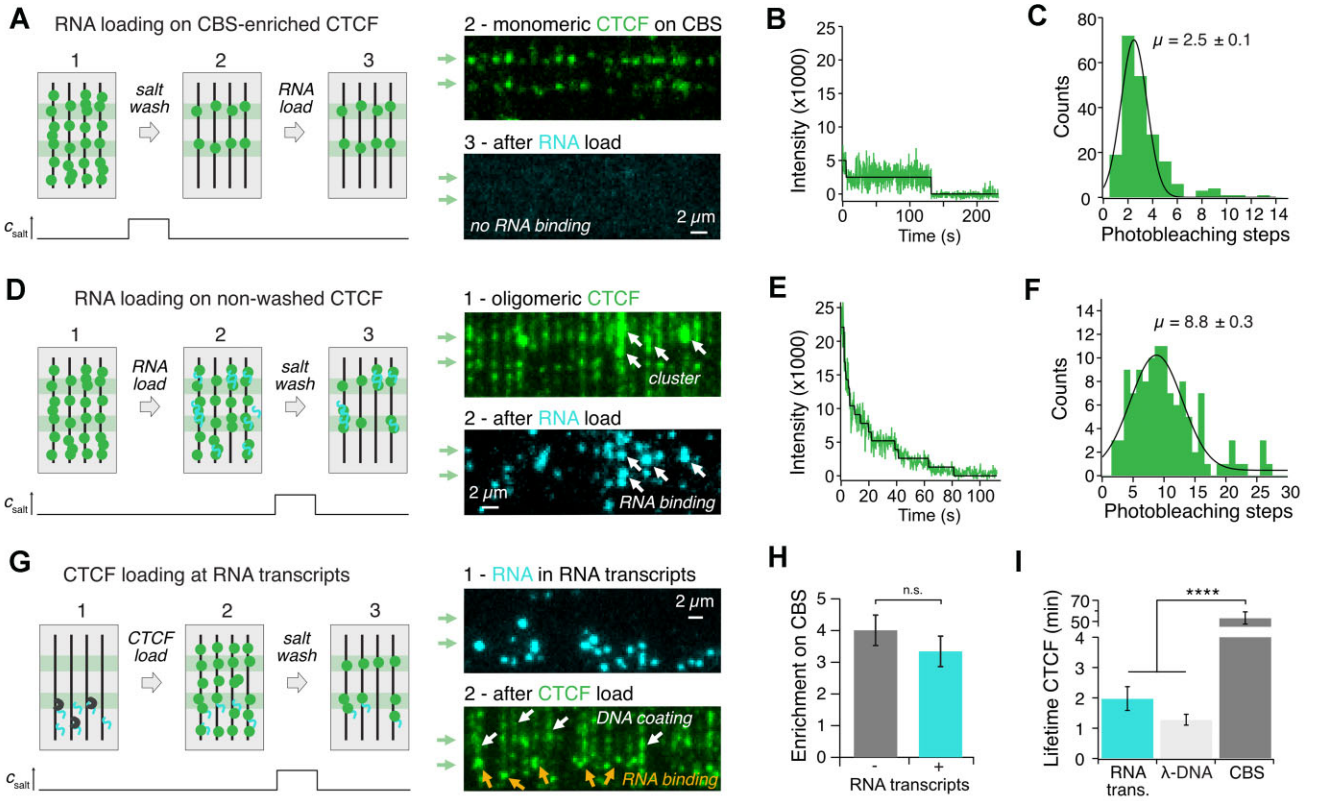


Figure 5. RNA recruitment by oligomeric CTCF. (**A**) 100 bp Cy3-UTP labeled RNA-loading (25 ng/µl) on CBS-enriched Alexa-Fluor 660-CTCF. Left: Experimental workflow. Right: TIRF microscopy at 100 ms illumination time of CBS-bound CTCF (green) before RNA load (top) and of Cy3-UTP labeled RNA (cyan) after RNA load (bottom). No RNA capture was observed. (**B**) Representative intensity curve of a two-step CTCF photobleaching event. Bleaching steps are illustrated by a black line. (**C**) Histogram of CTCF bleaching steps on 4× CBSs. Same as Supplementary Figure S1C, added here for comparison. (**D**) 100 bp Cy3-UTP labeled RNA-loading (25 ng/µl) on clustered Alexa-Fluor 660-CTCF. Left: Experimental workflow. Right: TIRF microscopy at 100 ms illumination time of CTCF clusters (green) before RNA load (top) and of Cy3-UTP labeled RNA (cyan) after RNA load. RNA is recruited to CTCF clusters (white arrows). (**E**) Representative intensity curve of multi-step CTCF bleaching event of a CTCF-RNA cluster. Steps are illustrated by a black line. (**F**) Histogram of CTCF bleaching steps in CTCF-RNA clusters. (**G**) 10 nM ATTO-643-CTCF loading at RNA transcripts. Left: Experimental workflow. RNA transcripts were formed by loading T7-Pol and facilitating transcription as before. Right: TIRF microscopy at 100 ms illumination time of Cy3-UTP labeled RNA in RNA transcripts (cyan) before CTCF load (top) and of CTCF (green) after loading (bottom). CTCF is partially recognizing RNA transcripts (orange arrows), but is mainly coating the DNA (white arrows). (**H**) Enrichment of CTCF on 4× CBSs is not significantly different in presence or absence of RNA transcripts. (**I**) CTCF lifetime on RNA transcripts is similar to lifetimes on λ-DNA and significantly smaller than on 4× CBSs.

showed that CTCF binds stably on CBSs, while it displays diffusion and low lifetimes on unspecific DNA (Figure 1C–E) (11,12). By performing measurements at different laser frame delay and correcting photobleaching data, we showed that CTCF has a much higher lifetime on CBSs than previously observed *in vitro* (11,12) and also *in vivo* (45,63–67). The lower lifetime *in vivo* might result from interactions with other chromatin bound proteins, as CTCF in resting B-cells showed a lifetime similar to our results (65). In our transcription assay, CTCF and CTCF-SA were pushed off CBSs by a transcribing polymerase (Figure 4D-F), which supports the theory of chromatin-bound proteins reducing CTCF's lifetime on DNA, an effect which might be weaker in resting B-cells, where transcription elongation is repressed (65). Our data is therefore consistent with a model where transcription impacts CTCF's residence time on chromatin. Transcription can displace cohesin *in vitro* (68), disrupt its localization at CTCF sites (69) and act as a mobile loop extrusion barrier (70). CTCF anchors near active genes might therefore be less stringently positioned and cause more diffuse interactions (71).

Correspondingly, we also observed CTCF to have an impact on transcription. Single-molecule transcription assays revealed a higher amount of impaired transcription and a slower transcription rate in the presence of CTCF (Figure 4F–H). A study on a bacterial RNA-Pol has shown that the length of transcription pauses, and the appearance of backtracking events is increased by transcription-opposing forces created by the nascent RNA (72). We speculate that CTCF, once pushed off its CBS by the polymerase, might be evicted from DNA by the nascent RNA (73,74) and change the secondary structure of the transcript. Permanently paused states could also be caused by polymerase unbinding and CTCF capturing the RNA chain. Transcription could therefore enable CTCF to recruit RNA to CBSs and lead to an increased insulation at domain boundaries (75). Impaired transcription by CTCF could also regulate Pol II pausing (55), which has been linked to alternative splicing (35).

A different explanation for transcription impairment is CTCF clustering and consequential bridging of multiple DNAs or RNAs by CTCF clusters (26,44,76) causing an increased number of T7 collisions, which can lead to T7 displacement from DNA (58). CTCF clustering has been linked to RNA binding (26,28), but it remains elusive whether CTCF's interaction with RNA leads to the formation of clus-

ters or if clustering is required for RNA capture. Here, we show that DNA-bound CTCF forms clusters, and that only clustered CTCF can capture RNA (Figure 5A–F). Oligomerization therefore enables RNA recruitment by chromatin-bound CTCF. Photobleaching experiments revealed an average number of 17 CTCF molecules per cluster, which is close to values observed in the nucleus (30). However, cluster sizes might differ in living cells, as both cohesin and transcription can shape cluster formation, with the former facilitating and the latter disrupting it (30).

Our data show that RNA modulates CTCF binding positions and stability *in vitro*. In the absence of RNA, CTCF is only stably bound to high-affinity binding sites, while RNA binding leads to stable binding also at low-affinity sites (Supplementary Figure S5A,C). Correspondingly, *in vivo*, low-affinity CBSs flank regions of active chromatin and have been associated with transcriptional regulation, while high-affinity CBSs more often flank repressed chromatin and are associated with regulating genome architecture (19). CTCF on high-affinity CBSs is highly stable (17,18), and loop domains can persist for hours without energy input, requiring stable anchoring by CTCF (77), in agreement with our observed long lifetimes on CBSs. Binding to low-affinity binding sites is less persistent (19) and disrupted by transcription inhibition or RNA-binding deficient mutants (29). Our results therefore support a model where RNA transcripts are an important regulator of CTCF binding positions and stability. This means that RNA is a key regulator in CTCF function, allowing a dynamic modulation of CTCF binding near active genes, but is less critical for CTCF's role in genome architecture. The latter is facilitated by CTCF's high CBS-stability, independently of clustering and RNA binding.

Our data demonstrate that oligomerization and RNA capture is independent of the unstructured termini (Supplementary Figure S5D). On the contrary, the termini might even impede cluster formation (Supplementary Figure S2B). CTCF is assumed to bind to RNA via its ZFs 1 and 10 as well as a C-terminal RNA binding region (28,29). However, a more recent study shows a ncRNA to interact with ZF3–6 and impede CTCF binding to its genomic locus (78). Since CTCF can bind RNA with multiple ZFs, and we found no RNA capture for monomeric CBS-bound CTCF, we propose a model in which CTCF oligomerization creates unoccupied ZFs, enabling simultaneous RNA and DNA-binding. This would allow CTCF to create an RNA interaction hub on chromatin possibly involved in the formation of transcriptional condensates (31) and RNA-dependent recruitment of cohesin (79).

Both CTCF and RNA can stabilize SA-DNA binding independent of cohesin

Cohesin's subunit SA colocalizes with R-loops *in vivo*, possibly by a direct interaction with RNA, or with other RNA-binding proteins (23). Our experiments show that in the absence of RNA, SA preferentially interacts with AT-rich DNA-regions, with below-average occupancy on CBSs, fast binding dynamics and low salt stability (Figure 3F–I). In contrast, SA displayed stronger colocalization with and a significantly more stable binding at RNA transcripts than on DNA. (Supplementary Figure S5G–J). SA-RNA interactions could therefore facilitate their localization to R-loops and to active genes (23,80–82), as well as recruitment of cohesin-SA to CTCF sites (79). However, unlike CTCF, DNA-bound SA cannot recruit RNA. Instead SA is displaced from DNA by RNA (Supplementary Figure S5E-F), arguing that SA's binding site for RNA and DNA are identical, with higher affinity (Supplementary Figure S5J) (25) for RNA over DNA.

CTCF can directly recruit cohesin's subunit SA independently of RNA and other cohesin subunits, by significantly increasing lifetime and salt stability compared to DNA-bound SA (Figure 3I). Moreover, the CTCF termini are not required for interaction with SA, suggesting a ZF-mediated recruitment mechanism of SA to CBSs (Supplementary Figure S3A). SA enrichment is enhanced as it seems to be able to specifically recognize CTCF on CBSs (Figure 3A), most likely through recognizing different sequence-induced CTCF binding modes. CTCF is more salt-stable than any cohesin subunit on chromatin (83), and CTCF depletion leads to a decrease in cohesin residence time (84). We therefore propose that direct SA-ZF interactions increase the residence time of the complete cohesin-complex at TAD-borders, enabling the formation of long-lived loops (77). Two genome localization studies revealed distinct subpopulations of SA1-cohesin and SA2-cohesin (85,86). Interestingly the SA2-cohesin subpopulation displayed a comparatively lower colocalization with CTCF. Here we show that SA in the absence of CTCF is not stably associated to DNA at physiological salt concentrations (Figure 3G, I), in line with the SA2-cohesin subpopulation displaying a low residence time (86) and salt stability (85). In contrast, we show that SA-ZF interactions stabilize SA even above physiological salt concentrations on CBSs (Figure 3I). This could explain why the SA1-cohesin subpopulation in these studies, colocalized with CTCF at TAD borders, displayed a higher salt stability (85) and residence time (86). Our data therefore suggest that CTCF-mediated SA stabilization on CBSs can mediate different binding dynamics and roles of SA1-cohesin and SA2-cohesin (81,82,85). However, as we observe a similar stabilization of SA1 and SA2, it remains to be determined what other factors regulate the increased colocalization of CTCF with SA1-cohesin compared to SA2-cohesin *in vivo*.

Cohesin's Rad21 subunit was previously shown to be essential for the interaction between SA and the CTCF N-terminal SA recognition motif (21). However, another study showed that Rad21 is dispensable for an interaction between SA and the CTCF C-terminus (24). In our experiments, we did not observe any influence of Rad21 on CTCF-SA interactions, suggesting that SAs are stabilized by CTCF ZFs independently of other cohesin subunits. In agreement with this, Rad21-independent colocalization of CTCF and SA have also been observed *in vivo* (23). The direct recruitment of SA by CTCF ZFs as well as CTCF's high stability on CBSs and ability to perform secondary RNA capture could therefore regulate SA1 and SA2's different functions in transcription regulation and genome architecture (81,82,85), possibly also by cohesin-independent interactions (23).

Conclusion

CTCF's different binding modes on chromatin enable it to perform a variety of different functions inside the nucleus (Figure 6). When binding unspecifically to DNA, CTCF displays a low residence time. CTCF can form oligomers, which can perform secondary capture of RNA, leading to more stable DNA binding. These oligomers could therefore act as an interaction hub for additional proteins like cohesin's SA subunits

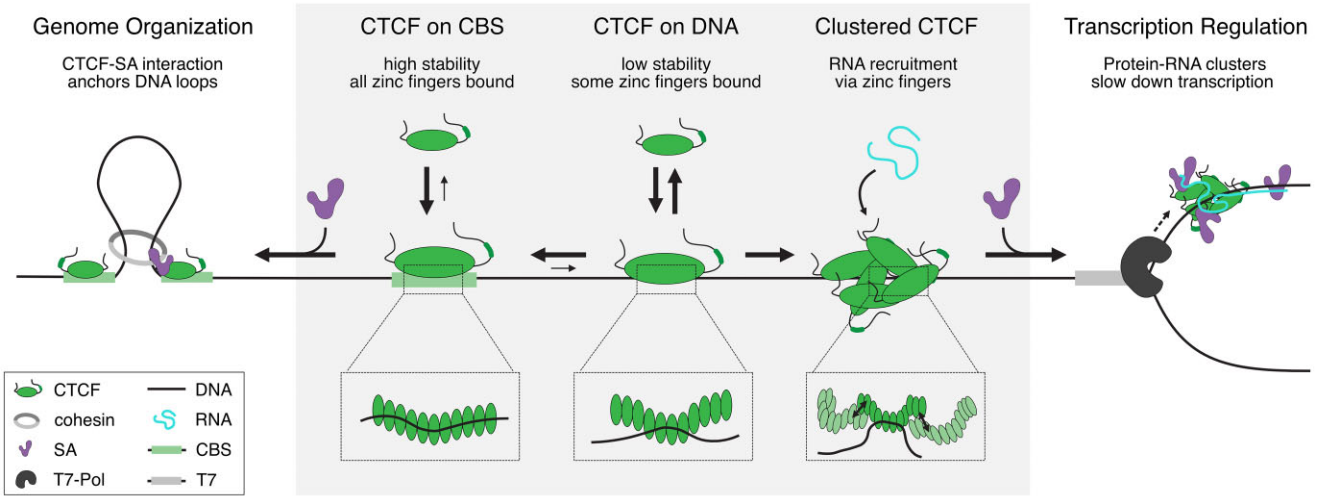


Figure 6. CTCF's nucleic acid interactions regulate diverse processes in the nucleus.

(79) or RNA Pol-II (31). Alternatively, monomeric CTCF can reach its target site by diffusion or rapid sampling of DNA, being stabilized on its CBS by more favorable ZF interactions. CBS-bound CTCF ZFs recruit SA by a Rad21-independent interaction, increasing SA's residence time. CTCF can therefore influence cohesin-independent functions of SAs in the nucleus (23). SAs may then recruit other cohesin subunits, facilitating TAD-formation. Alternatively, CTCF might block loop extrusion by its N-terminal interaction with cohesin (21) and subsequently increase loop-stability by the SA-ZF interaction. We conclude that the high-stability nucleic acid engagements of CTCF's multiple ZFs enable its diverse roles in transcription regulation and TAD-formation.

Data availability

The data that supports the findings of our study is available upon reasonable request.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Marvin Freitag, Jeanny Probst, Joelle Deplazes-Lauber and Sigrun Jaklin for preparation of λ-DNA constructs. Thanks also to Bingzhi Wang for preparing a Python script for single-molecule photobleaching analysis, and Daniel Bollschweiler for providing access to mass photometry. We thank Prof. Christof Gebhardt for providing the CTCF WT construct. Thanks also to all members of the Stigler lab for helpful discussions. We acknowledge support by the LMU Center for Nanoscience CeNS.

Author contributions: J.H. conducted the experiments. S.Z. conducted mass photometry experiments. N.-L.T. and J.H. established the single-molecule transcription assay. J.H. and J.S. analyzed the data. N.-L.T. analyzed CTCF diffusion data. All authors wrote the paper. J.S. designed the study, provided funding and supervised the project.

Funding

European Research Council starting grant [758124]. Funding for open access charge: European Research Council [758124].

Conflict of interest statement

None declared.

References

1. Lieberman-Aiden,E., Berkum,N.L., Van Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O., *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the Human Genome. *Science*, **326**, 289–294.
2. Rao,S.S.P., Huntley,M.H., Durand,N.C., Stamenova,E.K., Bochkov,I.D., Robinson,J.T., Sanborn,A.L., Machol,I., Omer,A.D., Lander,E.S., *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
3. Rowley,M.J. and Corces,V.G. (2018) Organizational principles of 3D genome architecture. *Nat. Rev. Genet.*, **13**, 1.
4. Hildebrand,E.M. and Dekker,J. (2020) Mechanisms and functions of chromosome compartmentalization. *Trends Biochem. Sci.*, **45**, 385–396.
5. Mirny,L.A., Imakaev,M. and Abdennur,N. (2019) Two major mechanisms of chromosome organization. *Curr. Opin. Cell Biol.*, **58**, 142–152.
6. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
7. Nora,E.P., Lajoie,B.R., Schulz,E.G., Giorgetti,L., Okamoto,I., Servant,N., Piolot,T., Van Berkum,N.L., Meisig,J., Sedat,J., *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381–385.
8. 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.
9. Wutz,G., Várnai,C., Nagasaka,K., Cisneros,D.A., Stocsits,R.R., Tang,W., Schoenfelder,S., Jessberger,G., Muhar,M., Hossain,M.J., *et al.* (2017) Topologically associating domains and chromatin

loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.*, **36**, 3573–3599.
10. Nora,E.P., Goloborodko,A., Valton,A.L., Gibcus,J.H., Uebersohn,A., Abdennur,N., Dekker,J., Mirny,L.A. and Bruneau,B.G. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, **169**, 930–944.
11. Davidson,I.F., Barth,R., Zaczek,M., van der Torre,J., Tang,W., Nagasaka,K., Janissen,R., Kerssemakers,J., Wutz,G., Dekker,C., *et al.* (2023) CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature*, **616**, 822–827.
12. Zhang,H., Shi,Z., Banigan,E.J., Kim,Y., Yu,H., Bai,X.-C. and Finkelstein,I.J. (2023) CTCF and R-loops are boundaries of cohesin-mediated DNA looping. *Mol. Cell*, **83**, 2856–2871.
13. Martinez,S.R. and Miranda,J.L. (2010) CTCF terminal segments are unstructured. *Protein Sci.*, **19**, 1110–1116.
14. Yin,M., Wang,J., Wang,M., Li,X., Zhang,M., Wu,Q. and Wang,Y. (2017) Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res.*, **27**, 1365–1377.
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. Yang,J., Horton,J.R., Liu,B., Corces,V.G., Zhang,X., Cheng,X. and Blumenthal,R.M. (2023) Structures of CTCF – DNA complexes including all 11 zinc fingers. *Nucleic Acids Res.*, **51**, 8447–8462.
17. Khoury,A., Achinger-Kawecka,J., Bert,S.A., Smith,G.C., French,H.J., Luu,P.L., Peters,T.J., Du,Q., Parry,A.J., Valdes-Mora,F., *et al.* (2020) Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat. Commun.*, **11**, 54.
18. Luan,J., Xiang,G., Gómez-García,P.A., Tome,J.M., Zhang,Z., Vermunt,M.W., Zhang,H., Huang,A., Keller,C.A., Giardine,B.M., *et al.* (2021) Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. *Cell Rep.*, **34**, 108783.
19. Marina-Zárate,E., Rodríguez-Ronchel,A., Gómez,M.J., Sánchez-Cabo,F. and Ramiro,A.R. (2023) Low-affinity CTCF binding drives transcriptional regulation whereas high-affinity binding encompasses architectural functions. *iScience*, **26**, 106106.
20. Vietri Rudan,M., Barrington,C., Henderson,S., Ernst,C., Odom,D.T., Tanay,A. and Hadjur,S. (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.*, **10**, 1297–1309.
21. Li,Y., Haarhuis,J.H.I., Cacciatore,Á.S., Oldenkamp,R., Marjon,S., Ruiten,V., Willems,L., Teunissen,H., Muir,K.W., Wit,E.D., *et al.* (2020) The structural basis for cohesin-CTCF anchored loops. *Nature*, **578**, 472–476.
22. Nora,E.P., Caccianini,L., Fudenberg,G., So,K., Kameswaran,V., Nagle,A., Uebersohn,A., Hajj,B., Saux,A.L., Coulon,A., *et al.* (2020) Molecular basis of CTCF binding polarity in genome folding. *Nat. Commun.*, **11**, 5612.
23. Porter,H., Li,Y., Neguembor,M.V., Beltran,M., Varsally,W., Martin,L., Cornejo,M.T., Pezic,D., Bhamra,A., Surinova,S., *et al.* (2023) Cohesin-independent STAG proteins interact with RNA and localise to R-loops to promote complex loading. *eLife*, **12**. e79386.
24. Xiao,T., Wallace,J. and Felsenfeld,G. (2011) Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol. Cell. Biol.*, **31**, 2174–2183.
25. Pan,H., Jin,M., Ghadiyaram,A., Kaur,P., Miller,H.E., Ta,M., Liu,M., Fan,Y., Mahn,C., Gorthi,A., *et al.* (2020) Cohesin SA1 and SA2 are RNA binding proteins that localize to RNA containing regions on DNA. *Nucleic Acids Res.*, **48**, 5639–5655.
26. Saldaña-Meyer,R., González-Buendía,E., Guerrero,G., Narendra,V., Bonasio,R., Recillas-Targa,F., Reinberg,D., Gonza,E., Reinberg,D. and Bonasio,R. (2014) CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.*, **28**, 723–734.
27. Kung,J.T., Kesner,B., An,J.Y., Ahn,J.Y., Cifuentes-Rojas,C., Colognori,D., Jeon,Y., Szanto,A., delRosario,B.C., Pinter,S.F., *et al.* (2015) Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. *Mol. Cell*, **57**, 361–375.
28. Hansen,A.S., Hsieh,T.S., Cattoglio,C., Reinberg,D., Darzacq,X., Tjian,R., Reinberg,D., Darzacq,X. and Tjian,R. (2019) Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. *Mol. Cell*, **76**, 395–411.
29. Saldaña-Meyer,R., Rodriguez-Hernaez,J., Escobar,T., Nishana,M., Jácome-López,K., Nora,E.P., Bruneau,B.G., Tsirigos,A., Furlan-Magaril,M., Skok,J., *et al.* (2020) RNA interactions are essential for CTCF-mediated genome organization. *Mol. Cell*, **76**, 412–422.
30. Gu,B., Comerci,C.J., McCarthy,D.G., Saurabh,S., Moerner,W.E. and Wysocka,J. (2020) Opposing effects of cohesin and transcription on CTCF organization revealed by super-resolution imaging. *Mol. Cell*, **80**, 699–711.
31. Lee,R., Kang,M., Kim,Y., Yang,B., Shim,H., Kim,S., Yang,C.M., Min,B., Jung,W., Lee,E., Joo,J., Park,G., Cho,W. and Kim,H. (2022) CTCF-mediated chromatin looping provides a topological framework for the formation of phase-separated transcriptional condensates. *Nucleic. Acids. Res.*, **50**, 207–226.
32. Kim,S., Yu,N.K. and Kaang,B.K. (2015) CTCF as a multifunctional protein in genome regulation and gene expression. *Exp. Mol. Med.*, **47**, e166.
33. Peña-Hernández,R., Marques,M., Hilmi,K., Zhao,T., Saad,A., Alaoui-Jamali,M.A., Rincon,S.V.D., Ashworth,T., Roy,A.L., Emerson,B.M., *et al.* (2015) Genome-wide targeting of the epigenetic regulatory protein ctcf to gene promoters by the transcription factor TFII-I. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E677–E686.
34. Chernukhin,I., Shamsuddin,S., Kang,S.Y., Bergström,R., Kwon,Y.-W., Yu,W., Whitehead,J., Mukhopadhyay,R., Docquier,F., Farrar,D., *et al.* (2007) CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol. Cell. Biol.*, **27**, 1631–1648.
35. Shukla,S., Kavak,E., Gregory,M., Imashimizu,M., Shutinoski,B., Kashlev,M., Oberdoerffer,P., Sandberg,R. and Oberdoerffer,S. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, **479**, 74–79.
36. Greene,E.C., Wind,S., Fazio,T., Gorman,J. and Visnapuu,M.L. (2010) DNA curtains for high-throughput single-molecule optical imaging. In *Meth. Enzymol.*, **472**, 293–315.
37. Tsekouras,K., Custer,T.C., Jashnsaz,H., Walter,N.G. and Pressé,S. (2016) A novel method to accurately locate and count large numbers of steps by photobleaching. *Mol. Biol. Cell*, **27**, 3601–3615.
38. E.L. Kaplan,P.M. (1958) Nonparametric estimation from incomplete observations. *Am. Stat. Assoc.*, **53**, 457–481.
39. Holden,S.J., Uphoff,S. and Kapanidis,A.N. (2011) DAOSTORM: an algorithm for high-density super-resolution microscopy. *Nat. Methods*, **8**, 279–280.
40. Oh,S., Russell,S. and Sastry,S. (2004) Markov chain Monte Carlo data association for general multiple-target tracking problems. *Proc. IEEE Conf. Decis. Control*, **1**, 735–742.
41. Vestergaard,C.L., Blainey,P.C. and Flyvbjerg,H. (2014) Optimal estimation of diffusion coefficients from single-particle trajectories. *Phys. Rev. E.*, **89**, 022726.
42. Chang,L.H., Ghosh,S., Papale,A., Luppino,J.M., Miranda,M., Piras,V., Degrouard,J., Edouard,J., Poncelet,M., Lecouvreur,N., *et al.* (2023) Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and Topologically Associating Domain boundaries. *Nat. Commun.*, **14**, 5615.
43. Kentepozidou,E., Aitken,S.J., Feig,C., Stefflova,K., Ibarra-Soria,X., Odom,D.T., Roller,M. and Flicek,P. (2020) Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.*, **21**, 5.
44. Zhou,R., Tian,K., Huang,J., Duan,W., Fu,H., Feng,Y., Wang,H., Jiang,Y., Li,Y., Wang,R., *et al.* (2022) CTCF DNA-binding domain

undergoes dynamic and selective protein–protein interactions. *iScience*, **25**, 105011.
45. Nakahashi,H., Kwon,K.K., Resch,W., Vian,L., Stavreva,D., Hakim,O., Pruett,N., Nelson,S., Qian,J., Dubois,W., *et al.* (2014) A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.*, **3**, 1678–1689.
46. Kim,T.H., Abdullaev,Z.K., Smith,A.D., Ching,K.a.,I.,D., Green,R.D., Zhang,M.Q., Lobanenkov,V.V. and Ren,B. (2008) Analysis of the vertebrate insulator protein CTCF binding sites in the human genome. *Cell*, **128**, 1231–1245.
47. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
48. Oguey,C., Foloppe,N. and Hartmann,B. (2010) Understanding the sequence-dependence of DNA groove dimensions: implications for DNA interactions. *PLoS One*, **5**, e15931.
49. 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.
50. Pugacheva,E.M., Kubo,N., Loukinov,D., Kang,S. and Kovalchuk,A.L. (2020) CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 2020–2031.
51. 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.
52. Yusufzai,T.M., Tagami,H., Nakatani,Y. and Felsenfeld,G. (2004) CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell*, **13**, 291–298.
53. Kang,H. and Lieberman,P.M. (2011) Mechanism of glycyrrhizic acid inhibition of Kaposi's sarcoma-associated herpesvirus: disruption of CTCF-cohesin-mediated RNA polymerase II pausing and sister chromatid cohesion. *J. Virol.*, **85**, 11159–11169.
54. Marina,R.J., Sturgill,D., Bailly,M.A., Thenoz,M., Varma,G., Prigge,M.F., Nanan,K.K., Shukla,S., Haque,N. and Oberdoerffer,S. (2016) TET-catalyzed oxidation of intragenic alternative splicing. *EMBO J.*, **35**, 335–355.
55. Paredes,S.H., Melgar,M.F. and Sethupathy,P. (2013) Promoter-proximal CCCTC-factor binding is associated with an increase in the transcriptional pausing index. *Bioinformatics*, **29**, 1485–1487.
56. Thomen,P., Lopez,P.J., Bockelmann,U., Guillerez,J., Dreyfus,M. and Heslot,F. (2008) T7 RNA polymerase studied by force measurements varying cofactor concentration. *Biophys. J.*, **95**, 2423–2433.
57. Galburt,E.A., Grill,S.W., Wiedmann,A., Lubkowska,L., Choy,J., Nogales,E., Kashlev,M. and Bustamante,C. (2007) Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature*, **446**, 820–823.
58. Zhou,Y. and Martin,C.T. (2006) Observed instability of T7 RNA polymerase elongation complexes can be dominated by collision-induced 'bumping.'. *J. Biol. Chem.*, **281**, 24441–24448.
59. Hark,A.T., Schoenherr,C.J., Katz,D.J., Ingram,R.S., Levorse,J.M. and Tilghman,S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, **405**, 486–489.
60. Bell,A.C. and Felsenfeld,G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 2–5.
61. Ruiz-Velasco,M., Kumar,M., Lai,M.C., Bhat,P., Solis-Pinson,A.B., Reyes,A., Kleinsorg,S., Noh,K.M., Gibson,T.J. and Zaugg,J.B. (2017) CTCF-mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. *Cell Syst.*, **5**, 628–637.
62. Vostrov,A.A., Taheny,M.J. and Quitschke,W.W. (2002) A region to the N-terminal side of the CTCF zinc finger domain is essential for activating transcription from the amyloid precursor protein promoter. *J. Biol. Chem.*, **277**, 1619–1627.
63. Agarwal,H., Reisser,M., Wortmann,C. and Gebhardt,J.C.M. (2017) Direct observation of cell-cycle-dependent interactions between CTCF and chromatin. *Biophys. J.*, **112**, 2051–2055.
64. Hansen,A.S., Pustova,I., Cattoglio,C., Tjian,R. and Darzacq,X. (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife*, **6**, e25776.
65. Kieffer-Kwon,K.R., Nimura,K., Rao,S.S.P., Xu,J., Jung,S., Pekowska,A., Dose,M., Stevens,E., Mathe,E., Dong,P., *et al.* (2017) Myc regulates chromatin decompaction and nuclear architecture during B cell activation. *Mol. Cell*, **67**, 566–578.
66. Soochit,W., Sleutels,F., Stik,G., Bartkuhn,M., Basu,S., Hernandez,S.C., Merzouk,S., Vidal,E., Boers,R., Boers,J., *et al.* (2021) CTCF chromatin residence time controls three-dimensional genome organization, gene expression and DNA methylation in pluripotent cells. *Nat. Cell Biol.*, **23**, 881–893.
67. Hansen,A.S., Amitai,A., Cattoglio,C., Tjian,R. and Darzacq,X. (2020) Guided nuclear exploration increases CTCF target search efficiency. *Nat. Chem. Biol.*, **16**, 257–266.
68. Davidson,I.F., Goetz,D., Zaczek,M.P., Molodtsov,M.I., Huis in 't Veld,P.J., Weissmann,F., Litos,G., Cisneros,D.A., Ocampo-Hafalla,M., Ladurner,R., *et al.* (2016) Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.*, **35**, 2671–2685.
69. Heinz,S., Texari,L., Hayes,M.G.B., Urbanowski,M., Chang,M.W., Givarkes,N., Rialdi,A., White,K.M., Albrecht,R.A., Pache,L., *et al.* (2018) Transcription elongation can affect genome 3D structure. *Cell*, **174**, 1522–1536.
70. Banigan,E.J., Tang,W., van den Berg,A.A., Stocsits,R.R., Wutz,G., Brandão,H.B., Busslinger,G.A. and Peters JM,L.A.M. (2023) Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc. Natl. Acad. Sci. U.S.A.*, **120**, e2210480120,
71. Harris,H.L., Gu,H., Olshansky,M., Wang,A., Farabella,I., Eliaz,Y., Kalluchi,A., Krishna,A., Jacobs,M., Cauer,G., *et al.* (2023) Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat. Commun.*, **14**, 3303.
72. Gabizon,R., Lee,A., Vahedian-Movahed,H., Ebright,R.H. and Bustamante,C.J. (2018) Pause sequences facilitate entry into long-lived paused states by reducing RNA polymerase transcription rates. *Nat. Commun.*, **9**, 2930.
73. Oh,H.J., Aguilar,R., Kesner,B., Lee,H.G., Kriz,A.J., Chu,H.P. and Lee,J.T. (2021) Jpx RNA regulates CTCF anchor site selection and formation of chromosome loops. *Cell*, **184**, 6157–6173.
74. Sun,S., Del Rosario,B.C., Szanto,A., Ogawa,Y., Jeon,Y. and Lee,J.T. (2013) Jpx RNA activates xist by evicting CTCF. *Cell*, **153**, 1537.
75. Islam,Z., Saravanan,B., Walavalkar,K., Farooq,U., Singh,A.K., Sabarinathan,R., Thakur,J., Pandit,A., Henikoff,S. and Notani,D. (2023) Active enhancers strengthen insulation by RNA-mediated CTCF binding at chromatin domain boundaries. *Genome Res.*, **33**, 1–17.
76. Mawhinney,M.T., Liu,R., Lu,F., Maksimoska,J., Damico,K., Marmorstein,R., Lieberman,P.M. and Urbanc,B. (2018) CTCF-induced circular DNA complexes observed by atomic force microscopy. *J. Mol. Biol.*, **430**, 759–776.
77. Vian,L., Pękowska,A., Rao,S.S.P., Kieffer-Kwon,K.R., Jung,S., Baranello,L., Huang,S.C., El Khattabi,L., Dose,M., Pruett,N., *et al.* (2018) The energetics and physiological impact of cohesin extrusion. *Cell*, **173**, 1165–1178.
78. Miyata,K., Imai,Y., Hori,S., Nishio,M., Loo,T.M., Okada,R., Yang,L., Nakadai,T., Maruyama,R., Fujii,R., *et al.* (2021) Pericentromeric noncoding RNA changes DNA binding of CTCF and inflammatory gene expression in senescence and cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2025647118.
79. Yao,H., Brick,K., Evrard,Y., Xiao,T., Camerini-Otero,R.D. and Felsenfeld,G. (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.*, **24**, 2543–2555.
80. Arruda,N.L., Carico,Z.M., Justice,M., Liu,Y.F., Zhou,J., Stefan,H.C. and Dowen,J.M. (2020) Distinct and overlapping roles

of STAG1 and STAG2 in cohesin localization and gene expression in embryonic stem cells. *Epigenetics Chromatin*, **13**, 32.
81. Casa,V., Gines,M.M., Gusmao,E.G., Slotman,J.A., Zirkel,A., Josipovic,N., Oole,E., Van Ijcken,W.F.J., Houtsmuller,A.B., Papantonis,A., *et al.* (2020) Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcriptional control. *Genome Res.*, **30**, 515–527.
82. Cuadrado,A., Giménez-Llorente,D., Kojic,A., Rodríguez-Corsino,M., Cuartero,Y., Martín-Serrano,G., Gómez-López,G., Marti-Renom,M.A. and Losada,A. (2019) Specific contributions of cohesin-SA1 and cohesin-SA2 to TADs and Polycomb domains in embryonic stem cells. *Cell Rep.*, **27**, 3500–3510.
83. Liu,Y. and Dekker,J. (2022) CTCF–CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity. *Nat. Cell Biol.*, **24**, 1516–1527.
84. Gabriele,M., Brandão,H.B., Grosse-Holz,S., Jha,A., Dailey,G.M., Cattoglio,C., Hsieh,T.H.S., Mirny,L., Zechner,C. and Hansen,A.S. (2022) Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, **376**, 476–501.
85. Kojic,A., Cuadrado,A., De Koninck,M., Giménez-Llorente,D., Rodríguez-Corsino,M., Gómez-López,G., Le Dily,F., Marti-Renom,M.A. and Losada,A. (2018) Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat. Struct. Mol. Biol.*, **25**, 496–504.
86. Wutz,G., Ladurner,R., Glenn,B., Hilaire,S., Stocsits,R.R., Ivanov,M.P., Schoenfelder,S., Lelij,P.V.D., Mechtler,K., Davidson,I.F., *et al.* (2020) ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin STAG1 from WAPL. *eLife*, **9**, e52091.

Received: January 3, 2024. **Revised:** March 21, 2024. **Editorial Decision:** April 22, 2024. **Accepted:** May 1, 2024