Molecular Cell, Volume *83*

Supplemental information

CTCF and R-loops are boundaries of cohesin-mediated DNA looping

Hongshan Zhang, Zhubing Shi, Edward J. Banigan, Yoori Kim, Hongtao Yu, Xiao-chen Bai, and Ilya J. Finkelstein

Molecular Cell, Volume *83*

Supplemental information

CTCF and R-loops are boundaries of cohesin-mediated DNA looping

Hongshan Zhang, Zhubing Shi, Edward J. Banigan, Yoori Kim, Hongtao Yu, Xiao-chen Bai, and Ilya J. Finkelstein

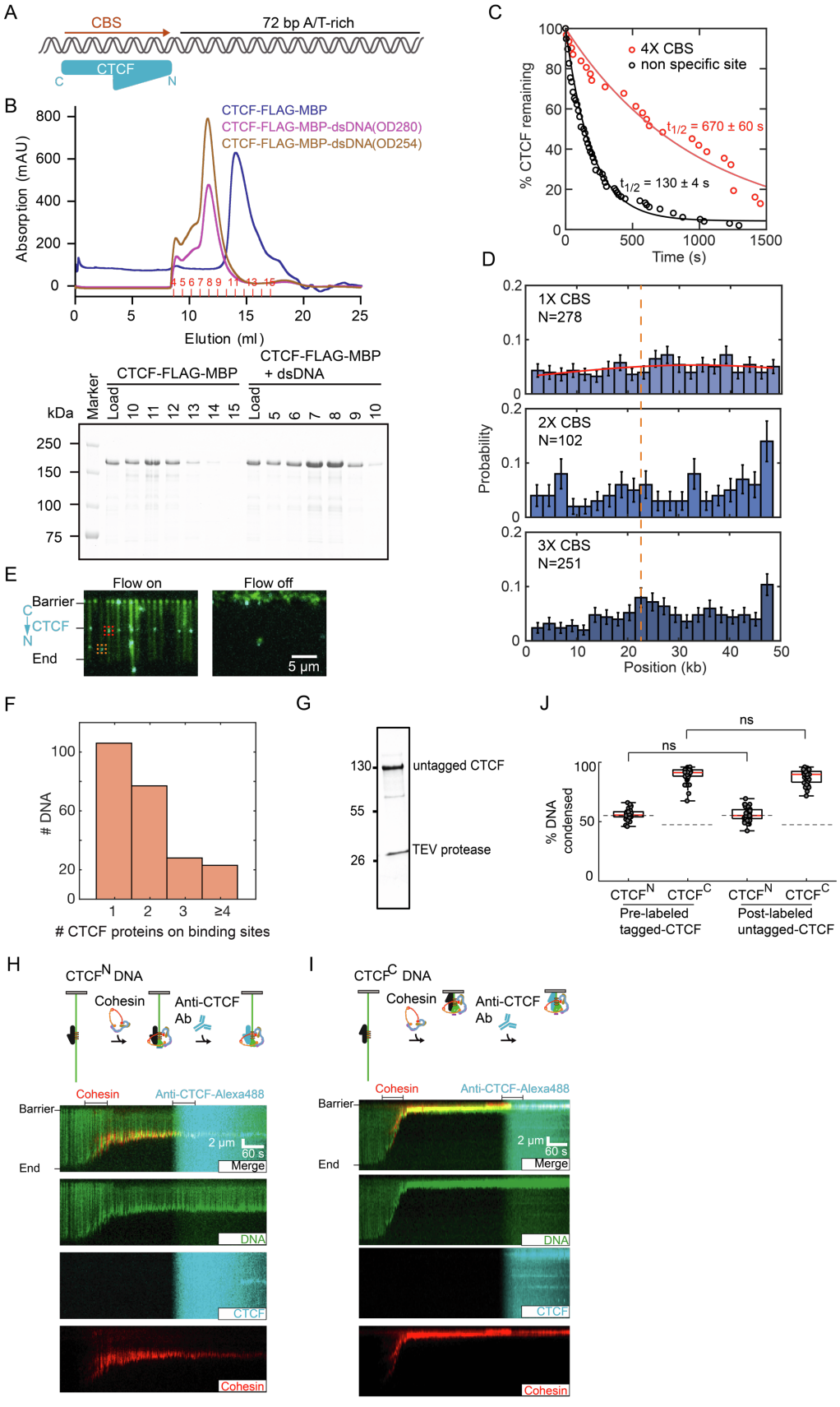
A
CBS
72 bp A/T-rich
CTCF
C
N
B
CTCF-FLAG-MBP
CTCF-FLAG-MBP-dsDNA(OD280)
CTCF-FLAG-MBP-dsDNA(OD254)
Absorption (mAU)
Elution (ml)
CTCF-FLAG-MBP
CTCF-FLAG-MBP + dsDNA
Marker
Load
kDa
250
150
100
75
C
4X CBS
non specific site
% CTCF remaining
t1/2 = 670 ± 60 s
t1/2 = 130 ± 4 s
Time (s)
D
1X CBS
N=278
2X CBS
N=102
3X CBS
N=251
Probability
Position (kb)
E
Flow on
Flow off
Barrier
CTCF
End
5 μm
F
DNA
CTCF proteins on binding sites
G
untagged CTCF
TEV protease
130
55
26
J
ns
% DNA condensed
CTCF^N
CTCF^C
Pre-labeled tagged-CTCF
Post-labeled untagged-CTCF
H
CTCF^N DNA
Cohesin
Anti-CTCF Ab
Anti-CTCF-Alexa488
Barrier
End
2 μm
60 s
Merge
DNA
CTCF
Cohesin
I
CTCF^C DNA

Figure S1. Purification and biochemical characterization of the CTCF-DNA complex. Related to Figure 1.

(A) Illustration of CTCF bound to the CTCF binding site (CBS).

(B) Size exclusion chromatography filtration profiles and corresponding SDS-PAGE analysis of CTCF and CTCF-DNA complexes.

(C) CTCF lifetimes on the 4X CBS DNA substrate (n=31) and on non-specific DNA sites (n=98). Solid lines are single exponential fits to the data. The half-life of each curve is indicated.

(D) CTCF binding distribution on DNA with single, double, and triple CBS. See Figure 1 for a binding distribution on the 4X CBS DNA substrate.

(E) We estimated the number of CTCF proteins on the 4x CBS DNA substrate by comparing the fluorescence intensity of the CTCF signal on the CBS (red box, n=233) to the average fluorescence intensity of CTCF patches on non-specific sites (orange box, n=237). Flow on/off indicates the specific and non-specific CTCF binding on DNA.

(F) The histogram showing the number of CTCF binding on 4x CBS DNA substrate.

(G) SDS-PAGE of wild type (wt) untagged CTCF.

(H-I) The experimental design (top) and kymographs (bottom) of single-molecule experiments with wt CTCF in the (H) $CTCF^{N}$ and (I) $CTCF^{C}$ orientations.

(J) Cohesin-driven DNA compaction is indistinguishable between wt and MBP-FLAG-tagged CTCF in either orientation. ns: no significance as determined by a t-test.

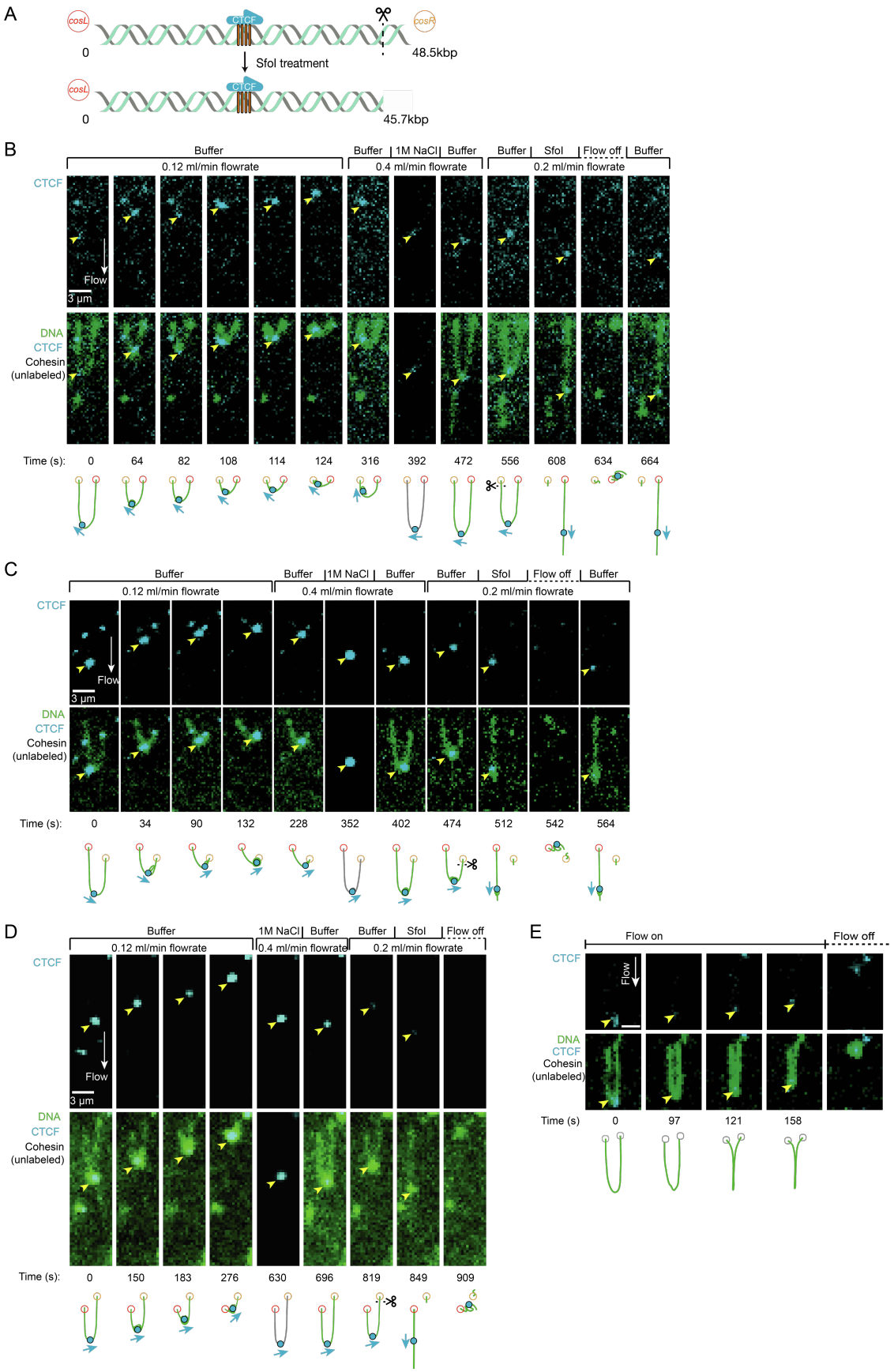
A
cosL
CTCF
cosR
0
48.5kbp
SfoI treatment
cosL
CTCF
0
45.7kbp
B
Buffer
0.12 ml/min flowrate
Buffer
1M NaCl
Buffer
0.4 ml/min flowrate
Buffer
SfoI
Flow off
Buffer
0.2 ml/min flowrate
CTCF
Flow
3 μm
DNA
CTCF
Cohesin
(unlabeled)
Time (s):
0
64
82
108
114
124
316
392
472
556
608
634
664
C
Buffer
0.12 ml/min flowrate
Buffer
1M NaCl
Buffer
0.4 ml/min flowrate
Buffer
SfoI
Flow off
Buffer
0.2 ml/min flowrate
CTCF
Flow
3 μm
DNA
CTCF
Cohesin
(unlabeled)
Time (s):
0
34
90
132
228
352
402
474
512
542
564
D
Buffer
0.12 ml/min flowrate
1M NaCl
Buffer
0.4 ml/min flowrate
Buffer
SfoI
Flow off
0.2 ml/min flowrate
CTCF
Flow
3 μm
DNA
CTCF
Cohesin
(unlabeled)
Time (s):
0
150
183
276
630
696
819
849
909
E
Flow on
Flow off
CTCF
Flow
DNA
CTCF
Cohesin
(unlabeled)
Time (s)
0
97
121
158

Figure S2. Cohesin compacts the segment between CTCF and *cosR* on U-shaped DNA. Related to Figure 1.

(A) The orientation of DNA is determined via optical restriction enzyme mapping with SfoI, which cleaves 1.2 kb away from *cosR*.

(B-C) Two examples of U-shaped DNA showing cohesin compacts the DNA segment on the N-terminal side of CTCF. Images are single frames taken at the indicated time of a single molecule movie.

(D) Example of cohesin compacting both segments of U-shaped DNA when it encounters $CTCF^{C}$. Both DNA ends are tethered to the flowcell surface. DNA is visualized with SYTOX Orange (green) and CTCF is labeled with an Alexa488-conjugated antibody (blue). DNA is compacted upon cohesin injection, which is disrupted by 1 M NaCl. Restriction enzyme mapping with SfoI, which cleaves near cosR, is used to identify the orientation of CTCF. Yellow arrows show the positions of CTCF.

(E) Cohesin bypasses CTCF during DNA compaction. CTCF (cyan) is at the CBS. The DNA (green) is tethered to the flowcell surface at both ends. Cohesin is injected and starts compacting the DNA at 121 s. CTCF remains at the CBS throughout cohesin translocation, likely indicating that cohesin is blocked by CTCF from one side. Yellow arrows show the CTCF positin. Scale bar: 3 μm.

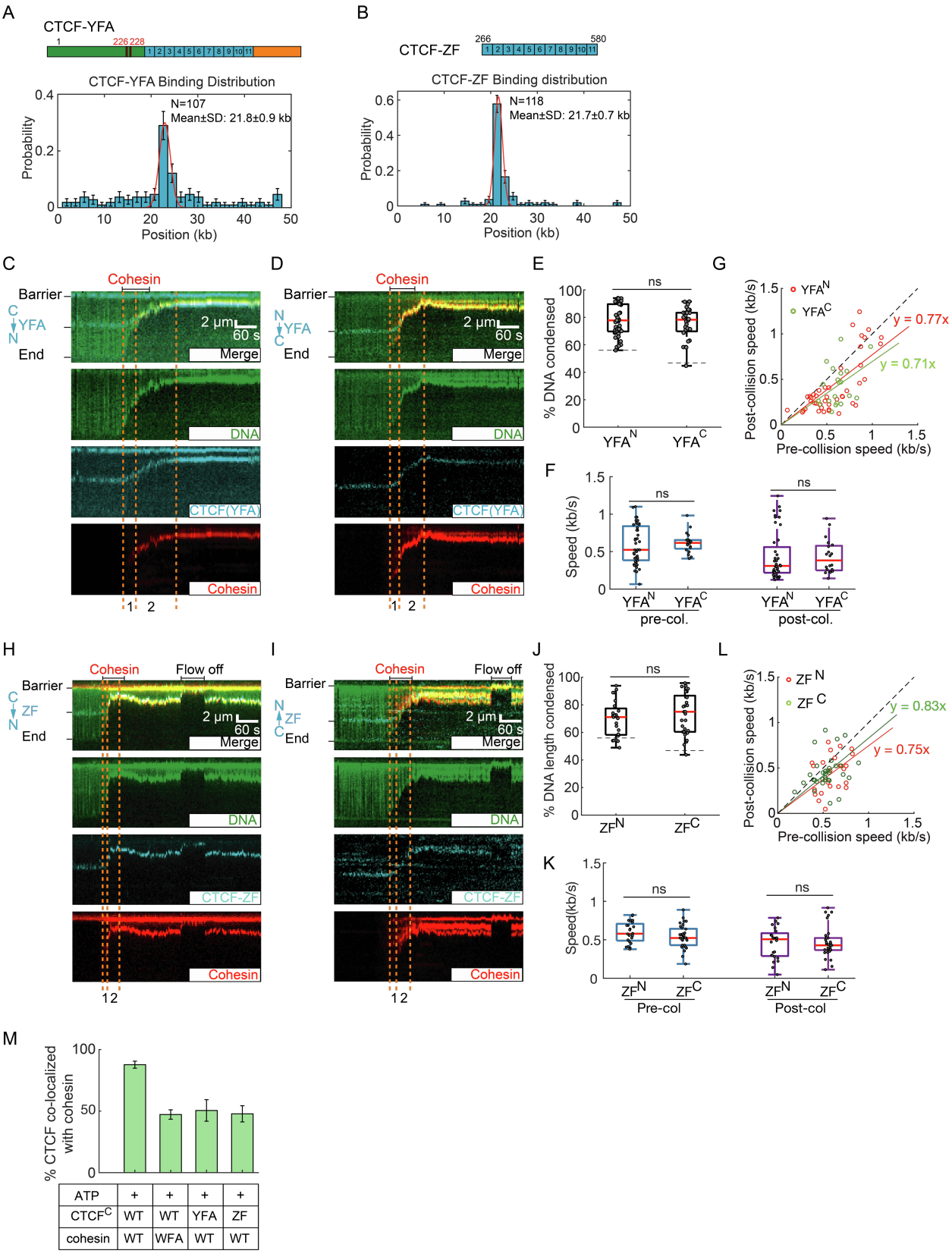
A
CTCF-YFA
1
226 228
CTCF-YFA Binding Distribution
N=107
Mean±SD: 21.8±0.9 kb
Probability
Position (kb)
B
CTCF-ZF
266
580
CTCF-ZF Binding distribution
N=118
Mean±SD: 21.7±0.7 kb
Probability
Position (kb)
C
Cohesin
Barrier
C
YFA
N
End
2 μm
60 s
Merge
DNA
CTCF(YFA)
Cohesin
1 2
D
Cohesin
Barrier
N
YFA
C
End
2 μm
60 s
Merge
DNA
CTCF(YFA)
Cohesin
1 2
E
ns
% DNA condensed
YFA^N
YFA^C
G
YFA^N
YFA^C
y = 0.77x
y = 0.71x
Post-collision speed (kb/s)
Pre-collision speed (kb/s)
F
ns
ns
Speed (kb/s)
YFA^N
YFA^C
pre-col.
post-col.
H
Cohesin
Flow off
Barrier
C
ZF
N
End
2 μm
60 s
Merge
DNA
CTCF-ZF
Cohesin
1 2
I
Cohesin
Flow off
Barrier
N
ZF
C
End
2 μm
60 s
Merge
DNA
CTCF-ZF
Cohesin
1 2
J
ns
% DNA length condensed
ZF^N
ZF^C
L
ZF^N
ZF^C
y = 0.83x
y = 0.75x
Post-collision speed (kb/s)
Pre-collision speed (kb/s)
K
ns
ns
Speed(kb/s)
ZF^N
ZF^C
Pre-col
Post-col
M
% CTCF co-localized with cohesin
ATP + + + +
CTCF^C WT WT YFA ZF
cohesin WT WFA WT WT

Figure S3. Characterization of CTCF mutants and their collision with cohesin. Related to Figure 2.

(A) CTCF-YFA and (B) CTCF-ZF binding distributions on DNA. Red curves: Gaussian fit. The center and S.D. of the fit are indicated in the histograms.

(C-D) Representative kymographs showing that cohesin compacts DNA after encountering $CTCF^{N}$-YFA and $CTCF^{C}$-YFA, respectively.

(E) Quantification of the compaction of DNA with $CTCF^{N}$-YFA (n=44) and $CTCF^{C}$-YFA (n=28).

(F) Cohesin speed pre- and post-collision with $CTCF^{N}$-YFA and $CTCF^{C}$-YFA. Dashed line: location of the CTCF binding sites on the DNA. Box plots indicate the median and quartiles.

(G) Pre- and post-collision speeds of individual cohesin molecules on DNA with $CTCF^{N}$-YFA (red) and $CTCF^{C}$-YFA (green). Dashed line is a slope of one to guide the eye.

(H-I) Representative kymographs showing that cohesin compacts $CTCF^{N}$-ZF and $CTCF^{C}$-ZF DNA, respectively.

(J) Quantification of the compaction of DNA with $CTCF^{N}$-ZF (n=24) and $CTCF^{C}$-ZF (n=33).

(K) The speed of cohesin translocation pre- and post-collision with $CTCF^{N}$-ZF and $CTCF^{C}$-ZF.

(L) Pre- and post-collision speeds of individual cohesin complexes on DNA with $CTCF^{N}$-ZF (red) or $CTCF^{C}$-YFA (green). Dashed line is a slope of one.

(M) The percent of CTCF variants co-localized with cohesin on $CTCF^{C}$ DNA (n>245 molecules for each condition from at least three flowcells).

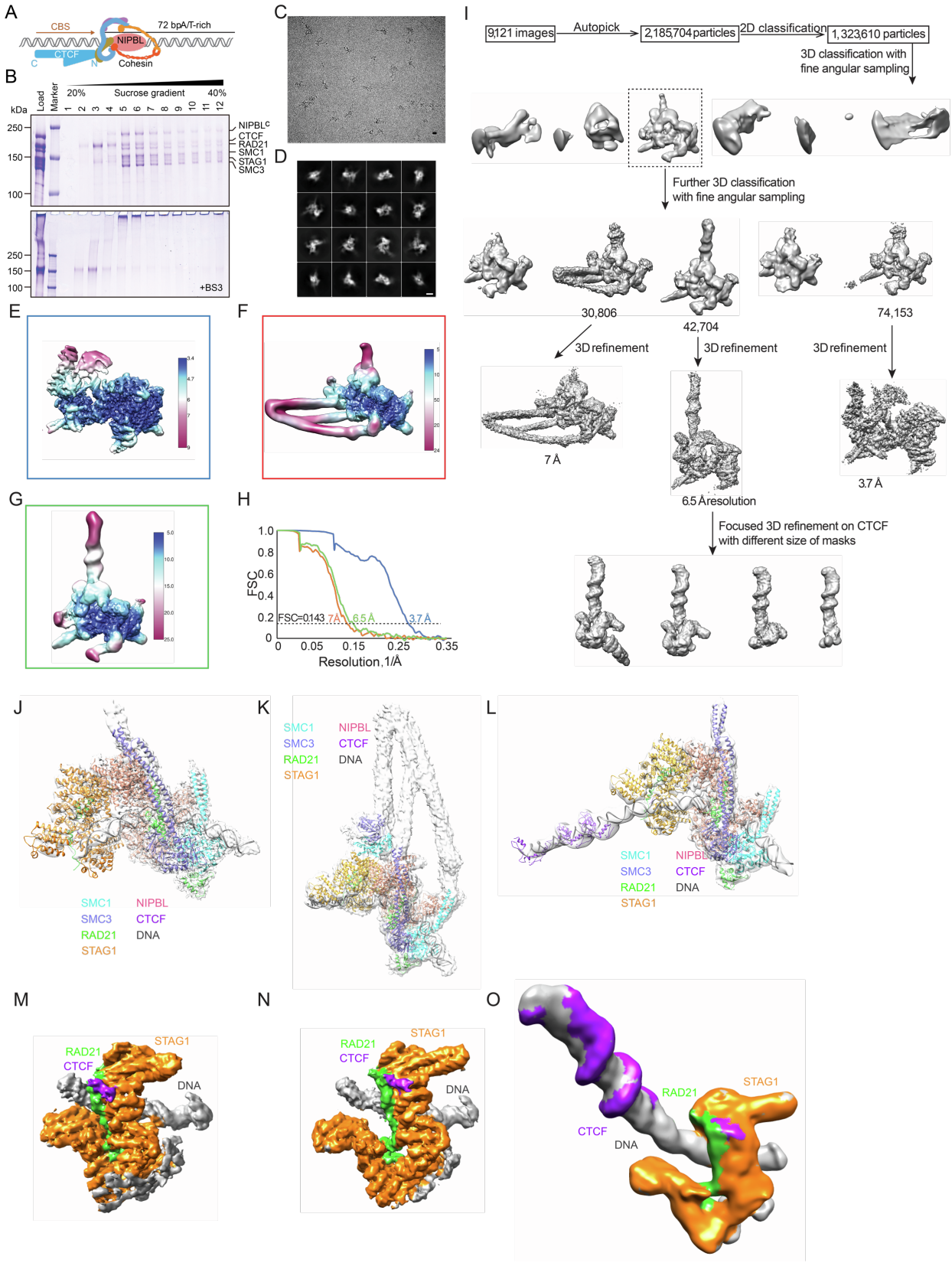
A
CBS
72 bpA/T-rich
NIPBL
CTCF
Cohesin
B
Load
Marker
20%
Sucrose gradient
40%
kDa
NIPBLC
CTCF
RAD21
SMC1
STAG1
SMC3
+BS3
C
D
E
F
G
H
FSC
Resolution,1/Å
FSC=0143
7Å
6.5Å
3.7Å
I
9,121 images
Autopick
2,185,704 particles
2D classification
1,323,610 particles
3D classification with fine angular sampling
Further 3D classification with fine angular sampling
30,806
42,704
74,153
3D refinement
3D refinement
3D refinement
7 Å
6.5 Å resolution
3.7 Å
Focused 3D refinement on CTCF with different size of masks
J
K
L
SMC1
SMC3
RAD21
STAG1
NIPBL
CTCF
DNA
M
N
O
STAG1
RAD21
CTCF
DNA

Figure S4. Cryo-EM sample preparation and structure determination of the human cohesin-NIPBL-CTCF-DNA complex, data processing workflow, models of cohesin, NIPBL and CTCF were docked to cryo-EM maps. Related to Figure 3.

(A) An illustration of cohesin, NIPBL and CTCF bound to the DNA substrate.

(B) SDS-PAGE analysis of the cohesin-NIPBL-CTCF-DNA complex with and without mild bis(sulfosuccinimidyl)suberate (BS3) crosslinking in the presence of ADP and BeF^{3-} after sucrose gradient ultracentrifugation.

(C) A representative micrograph of human cohesin-NIPBL-CTCF-DNA complex. Bar: 100 Å. (D) 2D class averages were selected for 3D reconstruction. Bar: 100 Å.

(E-G) Local resolution of the cryo-EM maps of the cohesin-NIPBL-DNA complex (E), the cohesin-NIPBL-DNA complex with folded coiled-coils (F), and the cohesin-NIPBL-CTCF-DNA complex (G). The box colors correspond to the curves in (H).

(H) Fourier shell correlation (FSC) curves of the cohesin-NIPBL-DNA complex (blue curve), the cohesin-NIPBL-DNA complex in folded state (orange curve), and the cohesin-NIPBL-CTCF-DNA complex (green curve).

(I) More than 2.1 million particles were picked from 9,121 images, and then were classified by 2D and 3D classification and refinement, generating three cryo-EM maps of the complex in distinct conformations.

(J) The cohesin-NIPBL-DNA complex at 3.7 Å resolution.

(K) The cohesin-NIPBL-DNA complex in folded state at 7.0 Å resolution.

(L) The cohesin-NIPBL-CTCF-DNA complex at 6.5 Å resolution. Maps and models are shown as transparent surfaces and cartoons, respectively.

(M) Locally refined maps of cohesin-NIPBL-DNA complex. The density of CTCF YDF motif was found on the surface of the STAG1-RAD21 subcomplex.

(N) Locally refined maps of cohesin-NIPBL-DNA complex in folded state.

(O) Locally refined maps of cohesin-NIPBL-CTCF-DNA complex.

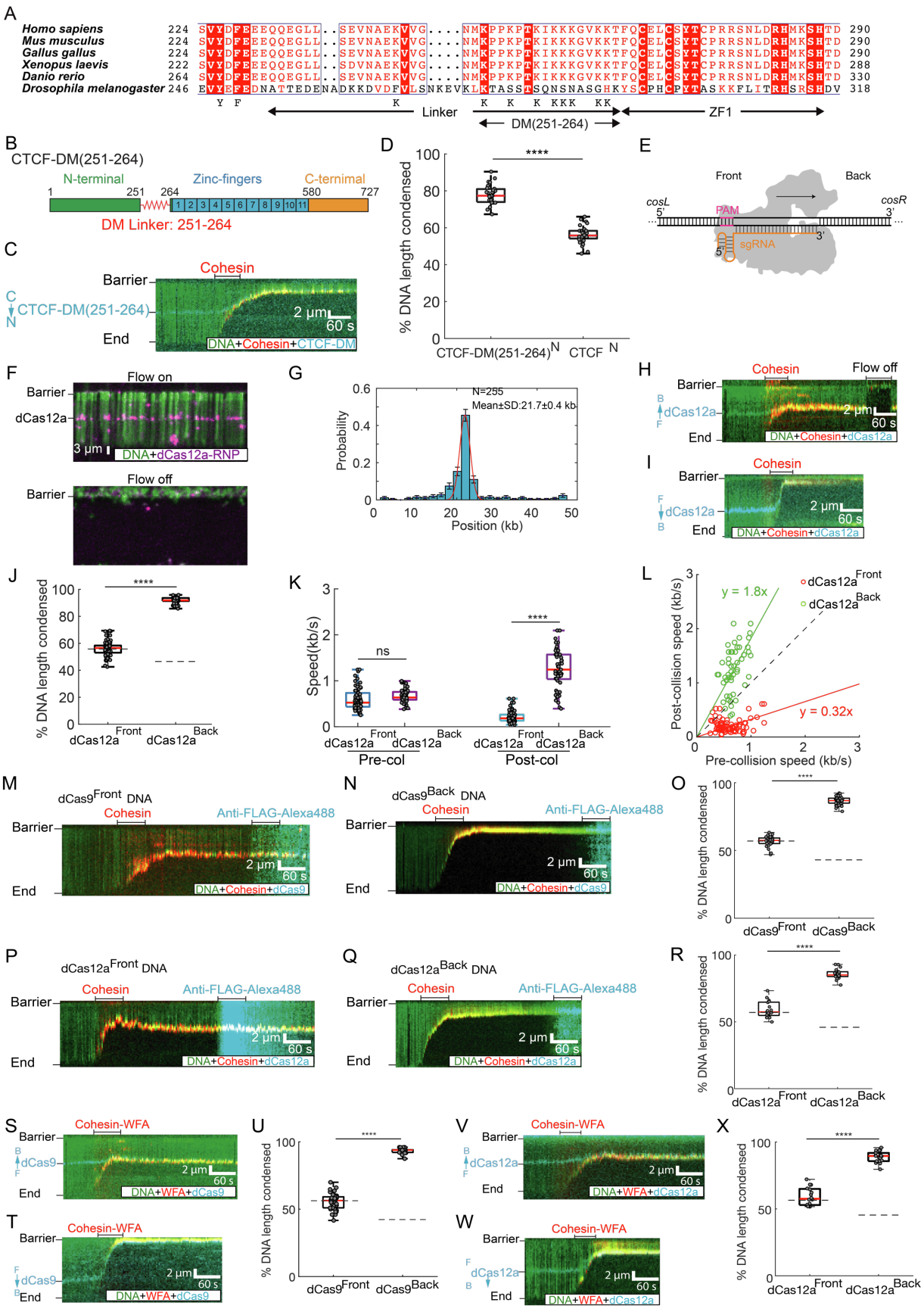
A
Homo sapiens
Mus musculus
Gallus gallus
Xenopus laevis
Danio rerio
Drosophila melanogaster
Linker
DM(251-264)
ZF1
B
CTCF-DM(251-264)
N-terminal
Zinc-fingers
C-ternimal
DM Linker: 251-264
C
Cohesin
Barrier
End
2 μm
60 s
DNA+Cohesin+CTCF-DM
D
% DNA length condensed
CTCF-DM(251-264)
CTCF
E
Front
Back
cosL
cosR
PAM
sgRNA
F
Flow on
Flow off
dCas12a
3 μm
DNA+dCas12a-RNP
G
N=255
Mean±SD:21.7±0.4 kb
Probability
Position (kb)
H
Flow off
DNA+Cohesin+dCas12a
I
J
K
Speed(kb/s)
ns
Pre-col
Post-col
L
y = 1.8x
y = 0.32x
Post-collision speed (kb/s)
Pre-collision speed (kb/s)
M
dCas9Front DNA
Anti-FLAG-Alexa488
DNA+Cohesin+dCas9
N
dCas9Back DNA
O
P
dCas12aFront DNA
Q
dCas12aBack DNA
R
S
Cohesin-WFA
DNA+WFA+dCas9
T
U
V
DNA+WFA+dCas12a
W
X

Figure S5. Sequence alignment of CTCF from the indicated species and related single-molecule experiments, related to Figure 3. Nuclease dead (d) Cas12a is a polar barrier to cohesin translocation. Related to Figure 4.

(A) The region between the YDF motif and zinc fingers is highly conserved from zebrafish to human but is divergent in fruit fly. The conserved YXF motif and conserved lysine residues are indicated below the diagram. *Drosophila melanogaster* linker amino acid 251-264 are indicated.

(B) Schematic of human CTCF with amino acid 251-264 replaced by the corresponding region of *Drosophila melanogaster*.

(C) Kymographs of cohesin collisions with CTCF-DM(251-264).

(D) Comparison of DNA compaction by the indicated CTCFs.

(E) Illustration of dCas12a binding on its target site. Orange: protospacer adjacent motif (PAM).

(F) Image of Alexa488-labled dCas12a on DNA with buffer flow on (top) and off (bottom). The position of the DNA target is indicated on the left.

(G) Binding distribution of dCas12a on DNA. Red curve: Gaussian fit.

(H-I) Representative kymographs showing that dCas12a^{Front} blocks cohesin (H), whereas dCas12a^{Back} accelerates cohesin translocation (I).

(J) Quantification of the condensed DNA in both dCas12a^{Front} and dCas12a^{Back} orientations (N>45 DNA molecules for each condition). Dashed line: dCas12a position on the DNA.

(K) Pre- and post-collision speeds for dCas12a^{Front} and dCas12a^{Back}.

(L) Correlation between the pre- and post-collision speeds of individual cohesin for the dCas12a^{Front} (red) and dCas12a^{Back} (green) orientations.

(M-N) Kymographs of cohesin-dCas9 collision on (M) dCas9^{Front} and (N) dCas9^{Back} DNA. In both experiments, dCas9 is fluorescently unlabeled during the collision and is labeled after the collision via anti-FLAG-Alexa488 antibodies.

(O) DNA compaction analysis shows no statistical difference between fluorescent and unlabeled dCas9 nucleases.

(P-Q) Kymographs of cohesin-dCas12a collision on (P) dCas12a^{Front} and (Q) dCas12a^{Back} DNA.

(R) DNA compaction by cohesin is indistinguishable between fluorescent and unlabeled dCas12a.

(S-T) Kymographs showing cohesin-WFA arrest by (S) dCas9^{Front}, but not dCas9^{Back} (T).

(U) Quantification of the condensed DNA in both orientations. Dashed line: dCas9 position on DNA.

(V-W) Representative kymographs showing (V) cohesin-WFA arrest by dCas12a^{Front}, but not dCas12a^{Back} (W).

(X) Quantification of the condensed DNA with both dCas12a orientations. Dashed line: dCas12a position on the DNA. t-test was used to determine the significant difference between experimental conditions. ****: $P < 0.0001$.

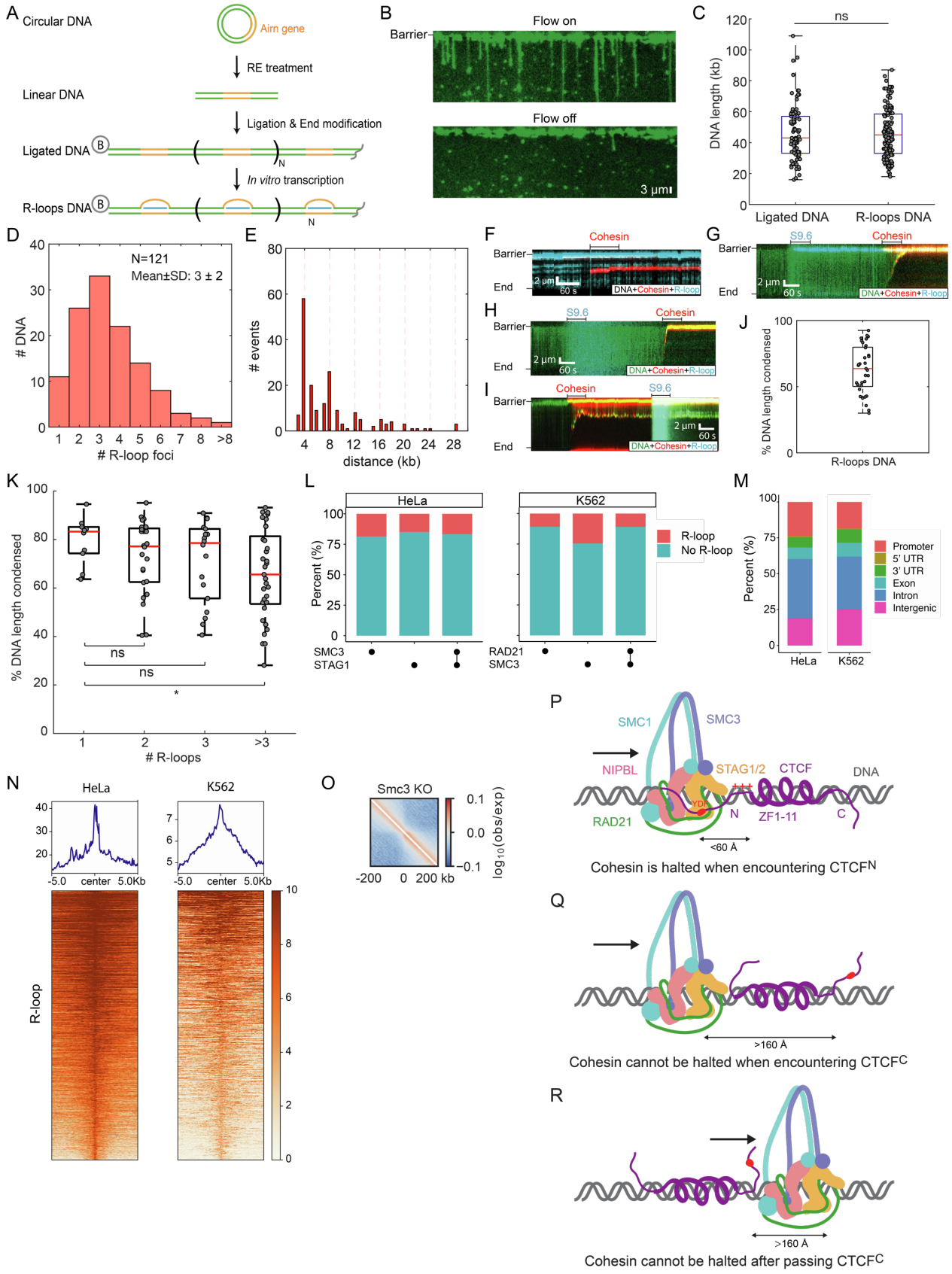
A
Circular DNA
Airn gene
RE treatment
Linear DNA
Ligation & End modification
Ligated DNA
In vitro transcription
R-loops DNA
B
Flow on
Barrier
Flow off
3 μm
C
ns
DNA length (kb)
Ligated DNA
R-loops DNA
D
N=121
Mean±SD: 3 ± 2
DNA
R-loop foci
E
events
distance (kb)
F
Cohesin
Barrier
End
2 μm
60 s
DNA+Cohesin+R-loop
G
S9.6
Cohesin
H
I
J
% DNA length condensed
R-loops DNA
K
% DNA length condensed
ns
*
R-loops
L
HeLa
K562
Percent (%)
R-loop
No R-loop
SMC3
STAG1
RAD21
M
Promoter
5' UTR
3' UTR
Exon
Intron
Intergenic
N
HeLa
K562
-5.0 center 5.0Kb
R-loop
O
Smc3 KO
-200 0 200 kb
log10(obs/exp)
P
SMC1
SMC3
NIPBL
STAG1/2
CTCF
DNA
RAD21
YDK
N
ZF1-11
C
<60 Å
Cohesin is halted when encountering CTCFN
Q
>160 Å
Cohesin cannot be halted when encountering CTCFC
R
>160 Å
Cohesin cannot be halted after passing CTCFC

Figure S6. Generation and characterization of R-loops DNA substrates, additional analysis of cohesin and R-loop collisions, and R-loops co-occur with cohesin in mammalian cell lines, related to Figure 5 and Figure 6.

(A) Overview of how R-loop DNA concatamers are generated. The *Airn* DNA sequence (orange) forms a stable R-loop when transcribed *in vitro*. This sequence is cloned into a 4 kb plasmid. The plasmid is linearized with a restriction enzyme (RE), ligated into concatamers, and transcribed with T3 RNA polymerase. RNase A treatment removes RNA that is outside the R-loop. B: biotin.

(B) Fluorescent image showing the distribution of DNA concatemer lengths after ligation. DNA is stained with SYTOX-Orange.

(C) The distribution of non-transcribed and transcribed ligated DNA (R-loops DNA) lengths. N>81 for both conditions.

(D) Number of R-loop foci per DNA molecule. The foci were stained with an Alexa488-conjugated S9.6 antibody.

(E) The distance between adjacent S9.6 antibody pairs (N=188). Dashed lines indicate multiples of 4 kb. At the experimental condition (0.12 mL min^{-1} flowrate and 100 nM SYTOX Orange in imaging buffer).

(F) Representative kymographs showing cohesin co-localized with R-loops.

(G) Representative kymograph showing that the S9.6 antibody doesn't stain any R-loops when the DNA is pre-treated with RNase H. Cohesin can also completely compact this substrate.

(H) Representative kymograph showing cohesin completely compacts non-transcribed ligated DNA.

(I) Representative kymograph showing that R-loops that are not labeled with S9.6 also slow cohesin. S9.6 was injected into the flowcell after cohesin translocation was complete.

(J) Quantification of percent DNA compacted for experiments where cohesin was injected into the flowcell prior to S9.6 injection, as shown in (E).

(K) DNA condensation as a function of the number of S9.6-stained R-loops. (n=10 for a s ingle R-loop; n=25 for tw o R-loop; n=18 for triple R -loops and n=38 for >3 R- loops). Red lines: median values.

(L) SMC3 and STAG1 positions overlap with R-loops in HeLa cells. SMC3 and RAD21 also overlap with R-loops in K562 cells.

(M) Genomic features of these overlapped peaks in HeLa and K562 cells.

(N) Read density profiles and heatmaps of R-loop reads across overlaps of SMC3, STAG1 and R-loop peaks in HeLa cells, and of RAD21, SMC3 and R-loop peaks in K562 cells.

(O) Map of average chromatin contact enrichment (observed-over-expected) in Smc3 KO MEFs centered on R-loops (n=39,680).

(P) CTCF YDF motif and positively charged linker region (+++) together halt cohesin translocation on DNA via interactions with the STAG1/2 subunit.

(Q-R) Cohesin can pass CTCF-bound CBSs either when encountering CTCF C-terminal side (Q) or crossing over CTCF (R). The linker between YDF and ZFs of CTCF is shorter than the dimension of either DNA-bound CTCF ZFs with linker or cohesin-NIPBL complex and therefore cannot reach its binding site in STAG1.

Table S1. Cryo-EM data collection and model statistics. Related to Figure 3.

Data collection and processing	
Magnification	46,296
Voltage (kV)	300
Electron exposure ($e^-/Å^2$)	60
Defocus rage (μm)	1.5 – 2.5
Pixel size (Å)	1.44
Symmetry imposed	*C*1
Initial particle images	2,185,704
Final particle images	42,704
Map resolution (Å)	6.5
FSC threshold	0.143
Refinement	
Initial model used (PDB code)	5T0U, 5YEL, 6WGE, 6QNX
Model resolution (Å)	5.8/6.2/7.1
FSC threshold	0/0.143/0.5
Model Composition	
Non-hydrogen atoms	34315
Proteins residues	3690
Nucleotides	216
Ligands	15
B factors ($Å^2$)	
Protein	240.41
Nucleotide	426.53
Ligand	178.68
Bonds RMSD	
Length (Å)	0.005
Angles (°)	0.929
Validation	
MolProbity score	2.09
Clash score	15.13
Ramachandran plot (%)	
Favored	93.87
Allowed	6.07
Outliers	0.05

Table S2. DNA oligos used in this work. Related to Figure 1-3.

Name	Description	Sequence
dsDNA_CBS	CTCF-DNA complex purification and protein structure study	GCAAGATTGCAGTGCCCACAGAGGC CAGCAGGGGGCGCTAGTGAGGTGGT TTTTATATGTTTTGTTATGTATTGTTT ATTTTCCCTTTAATTTTAGGATATGA AAACAAGAATTTATC The underlined sequence is the CTCF-binding site (CBS)
STAG1- W337A -F	Forward primer of PCR for STAG1 W337A	AAATACGTGGGCGCGACGCTGCACGATC GTCAGGGTG
STAG1- W337A -R	Reverse primer of PCR for STAG1 W337A	TCGTGCAGCGTCGCGCCCACGTATTTCA GATAGCTGTC
STAG1- F374A -F	Forward primer of PCR for STAG1 F374A	CTGAAATACGTGGGCGCGACGCTGCACG ATCG
STAG1- F374A -R	Reverse primer of PCR for STAG1 F374A	CGATCGTGCAGCGTCGCGCCCACGTATT TCAG
CTCF-1-F	Forward primer of PCR for CTCF	TATGGCCGGCCaATGGAAGGTGATGCAG TC
CTCF-727-A-NS	Reverse primer of PCR for CTCF	TTGGCGCGCCCCGGTCCATCATGCTGAG
CTCF_266-F	Forward primer of PCR for CTCF ZFs	TATGGCCGGCCaATGTTCCAGTGTGAGC TTTG
CTCF_580_NS-A	Reverse primer of PCR for CTCF ZFs	TTGGCGCGCCTGGGCCAGCACAATTATC
CTCF-Y226/F228A-F	Forward primer of PCR for CTCF Y226/F228A	GCAAAGATGTAGATGTGTCTGTCGCCGA TGCTGAGGAAGAACAGCAGGAGGGTC
CTCF-Y226/F228A-R	Reverse primer of PCR for CTCF Y226/F228A	GACCCTCCTGCTGTTCTTCCTCAGCATCG GCGACAGACACATCTACATCTTTGC
IF751	Forward primer of colony PCR for recombineering check	GAA CAA ACA ATA CCC AGA TTG CG
IF752	Reverse primer of colony PCR for recombineering check	GGA ATA TCT GGC GGT GCA AT
Lab06	Oligo for *cosR* end annealing of $CTCF^{C}$-DNA substrate	/5Phos/GGG CGG CGA CCT/3BioTEG

Lab07	Oligo for *cosL* end annealing of $CTCF^{N}$-DNA substrate	/5Phos/AGG TCG CCG CCC/3BioTEG
Lab08	Oligo for *cosL* end annealing of $CTCF^{C}$-DNA substrate	/5Phos/AGG TCG CCG CCC/3Dig_N
Lab09	Oligo for *cosR* end annealing of $CTCF^{N}$-DNA substrate	/5Phos/GGG CGG CGA CCT/3Dig_N

Table S3. Single guide RNAs used in this work. Related to Figure 4.

Name	RNA sequence
dCas9 sgRNA	GUG AUA AGU GGA AUG CCA UGG UUU UAG AGC UAG AAA UAG CAA GUU AAA AUA AGG CUA GUC CGU UAU CAA CUU GAA AAA GUG GCA CCG AGU CGG UGC UUU U
dCas12a crRNA	GUC AAA AGA CCU UUU UAA UUU CUA CUC UUG UAG AUA GGA UGA ACA GUU CUG GCU GGA GU

Table S4. The list of DRIP-Seq, ChIP-Seq and Hi-C data for Mouse Embryonic Fibroblasts. Related to Figure 6.

Mus musculus cell line	Seq type	GEO accession number	Pull-down in the indicated knockout (KO)	Citation
Mouse Embryonic Fibroblasts	DRIP-seq	GSE70189	R-loop	Sanz et al., 2016 [63]
	ChIP-seq	GSM1979724	Stag1	Busslinger et al., 2017 [62]
		GSM1979725	Stag1	
		GSM1979726	Stag1 in CTCF KO	
		GSM1979727	Stag1 in CTCF KO	
		GSM1979728	Stag1 in CTCF KO	
		GSM1979732	Stag1 in Wapl KO	
		GSM1979733	Stag1 in Wapl KO	
		GSM1979734	Stag1 in Wapl KO	
		GSM1979735	Stag1 in CTCF Wapl KO	
		GSM1979736	Stag1 in CTCF Wapl KO	
		GSM1979737	Rad21	
		GSM1979738	Rad22	
		GSM1979739	Rad21 in CTCF KO	
		GSM1979740	Rad21 in CTCF KO	
		GSM2221806	RAD21 in CTCF KO	
		GSM1979744	RAD21 in Wapl KO	
		GSM1979745	RAD21 in Wapl KO	
		GSM1979746	RAD21 in Wapl KO	
		GSM1979747	RAD21 in CTCF Wapl KO	
		GSM1979748	RAD21 in CTCF Wapl KO	
	Hi-C	GSE196621	Not applicable	Banigan et al., 2023 [20]

Table S5. Called peaks, peak overlaps and overlap p-values of ChIP/DRIP-seq for mouse embryonic fibroblasts. Related to Figure 6.

Mouse Embryonic Fibroblasts cell line	WT				CTCF KO				Wapl KO				DKO			
Genotype	Rad21	Stag1	Rad21 &Stag1	R-loop	Rad21	Stag1	Rad21 &Stag1	R-loop	Rad21	Stag1	Rad21 &Stag1	R-loop	Rad21	Stag1	Rad21 &Stag1	R-loop
# Total peaks	20891	27088	19952	35894	37713	34925	28419	35894	26346	27673	24070	35894	12841	12515	7923	35894
# Peaks overlapped with R-loop (at least 1 bp in common)	2920	3952	2695	NA	8114	9228	7037	NA	3826	4064	3356	NA	2699	3136	1654	NA
P-value (Fisher's exact test, bedtools)	< 1e-16	< 1e-16	< 1e-16	NA	< 1e-16	< 1e-16	< 1e-16	NA	< 1e-16	< 1e-16	< 1e-16	NA	< 1e-16	< 1e-16	< 1e-16	NA
P-value (ChiPseeker, nShuffle=10000)	< 1e-4	< 1e-4	< 1e-4	NA	< 1e-4	< 1e-4	< 1e-4	NA	< 1e-4	< 1e-4	< 1e-4	NA	< 1e-4	< 1e-4	< 1e-4	NA
P-value (Genomic HyperBrowser, Monte Carlo sampling=10000)	< 1e-4	< 1e-4	< 1e-4	NA	< 1e-4	< 1e-4	< 1e-4	NA	< 1e-4	< 1e-4	< 1e-4	NA	< 1e-4	< 1e-4	< 1e-4	NA

Note: We used the R-loop peaks called in WT cells across all KO experiments (shown in grey) because DRIP-seq data was not available for these knockouts.

Table S6. ChIP-Seq and DRIP-Seq sample list for human cell lines. Related to Figure S6.

Human cell line	Seq type	GEO accession number	Pull-down in the indicated knockout (KO)	Citation
HeLa	DRIP-seq	GSM2452072	R-loop	Hamperl et al., 2017 [66]
		GSM2668157	R-loop	
	ChIP-seq	GSM3619484	SMC3	Holzmann et al., 2019 [67]
		GSM3619485	EGFP-SA1	
K562	DRIP-seq	GSM1720619	R-loop	Sanz et al., 2016 [63]
	ChIP-seq	GSM935310	SMC3	Pope et al., 2014 [68]
		GSM935319	RAD21	

Table S7. Called peaks, peak overlaps and overlap p-values of ChIP/DRIP-seq for human cell lines. Related to Figure S6.

Human cell line	**HeLa**				**K562**			
Pull-downs	SMC3	SA1	SMC3 &SA1	R-loop	RAD21	SMC3	RAD21 &SMC3	R-loop
# Total peaks	29459	39799	26099	89246	12761	42740	12609	50482
# Peaks overlapped with R-loop	5470	5910	4335	NA	1381	10467	1374	NA
P-value (Fisher's exact test, bedtools)	< 1e-16	< 1e-16	< 1e-16	NA	< 1e-16	< 1e-16	< 1e-16	NA
P-value (ChIPseeker, nShuffle=10000)	< 1e-4	< 1e-4	< 1e-4	NA	< 1e-4	< 1e-4	< 1e-4	NA
P-value (Genomic HyperBrowser, Monte Carlo sampling=10000)	< 1e-4	< 1e-4	< 1e-4	NA	< 1e-4	< 1e-4	< 1e-4	NA