Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwigs-Maximilians-Universität München


# Analysing the various layers of genome architecture using a high-throughput single-molecule technique

Jonas Huber

aus

München, Deutschland

2025

**Erklärung**

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Johannes Stigler betreut.

**Eidesstaatliche Versicherung**

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 08.04.2025

_____

Jonas Huber

Dissertation eingereicht am: 16.01.2025

1 Gutachter: Professor Doktor Johannes Stigler

2 Gutachter: Professor Doktor Julian Stingele

Mündliche Prüfung am: 04.04.2025

This thesis has been prepared from April 2018 to January 2025 in the laboratory of Prof. Dr. Johannes Stigler at the Gene Center of the Ludwigs-Maximilians-Universität München, Germany.

This is a cumulative thesis based on the following publications:

**Huber,J.**, Tanasie,N.-L., Zernia,S. and Stigler,J. (2024) Single-molecule imaging reveals a direct role of CTCF's zinc fingers in SA interaction and cluster-dependent RNA recruitment. *Nucleic Acids Res.*, **52**, 6490–6506.

Gutierrez-Escribano,P., Newton,M.D., Llauró,A., **Huber,J.**, Tanasie,L., Davy,J., Aly,I., Aramayo,R., Montoya,A., Kramer,H., Stigler J., Rueda D.S., Aragon L. (2019) A conserved ATP- And Scc2/4-dependent activity for cohesin in tethering DNA molecules. *Sci. Adv.*, **5**, eaay6804.

Contents

# Abstract

In eukaryotes, genetic information is stored in mitochondria and the nucleus as long strands of deoxyribonucleic acid (DNA) (1). Nuclear DNA extends for roughly 6.3 million base pairs (bp) with a total length of about 2 meters (1). As the nucleus is only 10 µm in diameter (2), DNA needs to be highly compacted to fit inside, which is a challenging task for eukaryotic cells. DNA compaction and organisation are accomplished by a variety of different DNA-interacting proteins, which additionally regulate genome accessibility and, hence, gene expression in different cells, tissue types and organisms (3–7).

To study DNA-protein interactions a single-molecule assay based on total internal fluorescence microscopy called DNA curtains is applied (8). Here hundreds of recombinantly expressed and fluorescently labelled proteins can be visualised on parallelly aligned DNA molecules simultaneously, which will be explained in more detail in **Chapter I 2.**.

Polymerases bind to and translocate on DNA to copy genetic information during DNA replication and transcription, requiring accessible DNA (9). This is influenced by DNA sequence, other proteins and the activity of the polymerase itself, which I present in **Chapter II 1.** and **Chapter II 4.**.

In the nucleus, DNA is normally wrapped into nucleosomes, composed of 146 bp of DNA wound around an octameric protein complex (10). DNA in this conformation, called chromatin fibre, is compacted and less accessible to regulatory proteins (11). Accessibility depends on the positioning and spacing of nucleosomes and in **Chapter II 2.** I analyse nucleosome positioning depending on DNA sequence.

In addition to chromatin fibre formation, additional compaction is required to fit the whole DNA into the nucleus. In the next compaction step DNA is folded into spatial domains called topologically associating domains (TADs) by a process called loop extrusion (12–15). TADs are characterised by shorter three-dimensional (3D) genomic distances, increasing regulatory interactions within them while decreasing interactions with neighbouring regions (16, 17). They are formed by the ring-shaped cohesin complex, which generates DNA loops and the architectural protein CCCTC-binding factor (CTCF) residing on its genomic binding sites as an anchor point for these loops (18, 19). Cohesin additionally functions in cell division, preventing early separation of sister chromatids (20). **Chapter II 3.** shows that cohesin can form tethers between two DNAs and stable bridges on single DNA molecules, revealing potential mechanisms for holding sister chromatids together and for loop formation. CTCF's high stability on its DNA binding site and its ability to recruit secondary binding partners like cohesin's SA subunit and ribonucleic acid (RNA) are displayed in **Chapter II 4.**.

Chromatin domains further assemble into higher-order chromatin compartments, referred to as A- (transcriptionally active) or B- (transcriptionally inactive) compartments (21). Their formation depends on phase separation, a process in which the interaction of specific nucleic acids and proteins leads to the formation of dynamic phases segregated from the surrounding liquid (22–26). Phase-separating proteins often contain intrinsically disordered regions (IDRs) and nucleic acid-binding domains (27, 28), as is the case for CTCF (29, 30). **Chapters II 4. and 5.** display that CTCF forms oligomers under physiological conditions, enabling it to form clusters with DNA and capture RNA, which might create an interaction hub for phase-separating proteins.

This study reveals mechanistic insights into genome organisation involving transcription, chromatin formation and the different layers of genome architecture by studying protein-DNA interactions and higher-order complex formation on a single-molecule level. Additionally, it sheds light on the mutual interplay of these complex processes.

# I Introduction

## I 1. Biological background

One of the most fundamental and complex questions in cellular biology is how the enormous amount of genetic information (2 m of DNA (11, 24)) is both compacted into the small space of the nucleus and at the same time accessible to allow a variety of nuclear processes like replication (31), transcription (32), gene regulation (33) and DNA repair (34) to occur. To achieve a better understanding of these different cellular activities, genome architecture can be divided into different hierarchies, starting with a single strand of DNA and ending up with the division into chromatin territories, which is depicted in **Figure 1A**. A huge breakthrough in revealing these different layers was the development of chromosome conformation capture techniques, allowing probing 3D DNA-DNA interactions between non-neighbouring chromatin and displaying interaction frequencies in genome contact maps (2, 35, 36).

At the largest scale, chromosomes fold into distinct territories (37), with more frequent intra – than interchromosomal interactions (38), visible as squares corresponding to the size of each chromosome on genome contact maps (39) **(Figure 1B)**. Chromosome territories are divided into A- (transcriptionally active) and B- (transcriptionally inactive) compartments, forming long-range interactions with regions containing the same epigenetic features, leading to a checkboard-like interaction pattern (21, 39) **(Figure 1C)**. Within compartments, DNA is organised into TADs, defined by an increased frequency in 3D DNA-DNA interactions within them observed as squares on genome contacts maps and less frequent interactions with neighbouring regions (16, 19, 40) **(Figure 1D)**. Strong interactions where the two boundaries of each TAD meet are visible as dots in genome contact maps (16, 19, 40) **(Figure 1D)**. Lamina-associated domains (LADs) are also defined by increased DNA self-interactions but contain mostly transcriptionally inactive regions, more frequently found in B-compartments and are usually positioned close to the nuclear lamina (41–43). LADs and TADs are further subdivided into smaller compartmental domains and loops, which form small strong interaction points visible as dots on genome contact maps (40, 44) **(Figure 1E)**. At the smallest scale, DNA is wrapped around histone octamers, forming a compacted structure resembling “beads on a string” (45) **(Figure 1A)**. In the following chapters, I will describe the proteins involved in these different layers of genome architecture.

### 1.1 The chromatin fibre and nucleosome remodelling

#### *1.1.1 Nucleosome formation*

Nucleosomes are formed by 146 bp of DNA wrapping in 1.75 left-handed superhelical turns around a histone octamer containing two copies of histones H2A, H2B, H3 and H4 each (10, 46, 47) **(Figure 2A)**. H3 and H4 form a tetramer at the centre of the DNA, while histones H2A and H2B form two dimers closer to the ends of the DNA turns, and are therefore more accessible for interaction with regulatory proteins (10, 48, 49). The energy required for wrapping DNA comes from electrostatic interactions formed at roughly every 10 bp via main chain amide nitrogens and arginine and lysine side chains with the DNA phosphate backbone at positions where the DNA minor groove is facing towards the histone octamer (10, 50). In *human* cells, nucleosomes form at roughly every 200 bp, which is called nucleosome repeat length (NRL), which can however differ based on cell

type and the genomic region (51, 52). It further differs between organisms as NLRs are 175 bp in *C. elegans* (53) and 165 bp in *yeast* cells (54, 55).

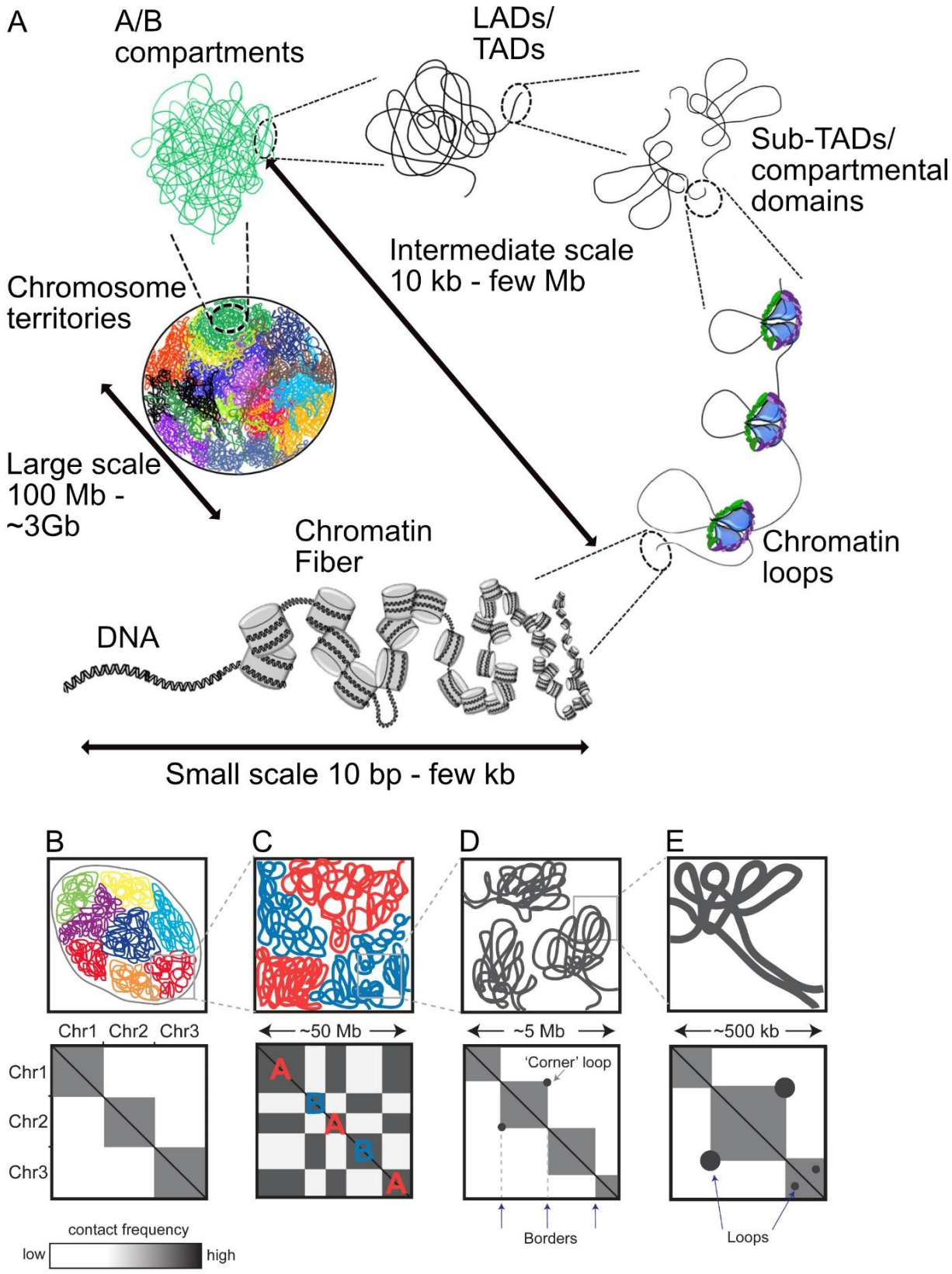



*Figure 1: Genome organisation across multiple scales. A) The DNA double helix is wrapping around histone octamers, forming the chromatin fibre. On the intermediate scale, DNA is further compacted by the formation of loops facilitated by genome architecture proteins, which leads to the creation of sub-TADs, TADs and LADs, with increased chromatin self-interactions within these domains.*

*Chromatin is further organised into segregated A/B compartments, differing in DNA accessibility and transcriptional activity. On the largest scale, there are chromosome territories defined by the position of each chromosome and more frequent intra- than interchromosomal interactions (38). Figure adapted from (37). B) Genome contact maps visualise DNA-DNA contact frequencies as a heatmap. Genomic positions are displayed on both x- and y-axis. The interactions between regions directly next to each other on the DNA strand therefore lead to a diagonal of high contact frequencies. More frequent intra- then interchromosomal interactions lead to additional squares of high contact frequency corresponding to the size of each chromosome. Figure from (39). C) A/B compartments form almost no contacts with neighbouring regions of the opposite type (A/B) but colocalise with distant compartments of the same epigenetic type leading to a patchwork-like pattern in genome contact maps. Figure from (39). D) Squares of high contact frequency along the diagonal visualise the shorter 3D DNA-DNA distances within TADs and increased distances to neighbouring TADs. TAD boundaries can be observed as dots of strong contact frequency between the two ends of each TAD. Figure from (39). E) Smaller loops within TADs are observed as small dots of high contact frequency. Figure from (39).*

Nucleosomes are not uniformly distributed and the exact positioning depends on different factors like DNA sequence (56–58), histone modifications and the interaction with other proteins (59). For genomic DNA nucleosome-bound sequences often display a characteristic recurrence of dinucleotides at every 10 bp (60–62) and the same periodicity can be observed in artificial sequences selection-optimised for high-affinity nucleosome formation like the 601 site *in vitro* (63). In contrast, AT-rich regions can destabilise histone-DNA interactions (64) and nucleosomes display a low occupancy on AT-rich regions in *yeast*, leading to the formation of nucleosome-free regions (NFRs) around transcription start sites (TSSs), which could increase accessibility for transcription factors (TFs) (54, 56, 65, 66). They are also depleted from GC-rich upstream regulatory regions (CpG islands) in mammals (51, 67, 68), which are instead often bound by TFs (68–73). Accurate positioning of nucleosomes might therefore also be involved in controlling genomic processes like transcriptional regulation (74).

### *1.1.2 Epigenetic features of histones*

Nucleosome positioning and therefore transcriptional regulation are further controlled by epigenetic modifications (75). Methylation of the C5-atom at CpG positions changes the DNA structure leading to a narrower minor groove, favouring the above-mentioned electrostatic interactions with arginine side chains (76, 77). This can lead to both stronger histone-DNA contacts and increased nucleosome occupancy, resulting in decreased DNA accessibility and transcriptional activity (78–80) **(Figure 2B)**.

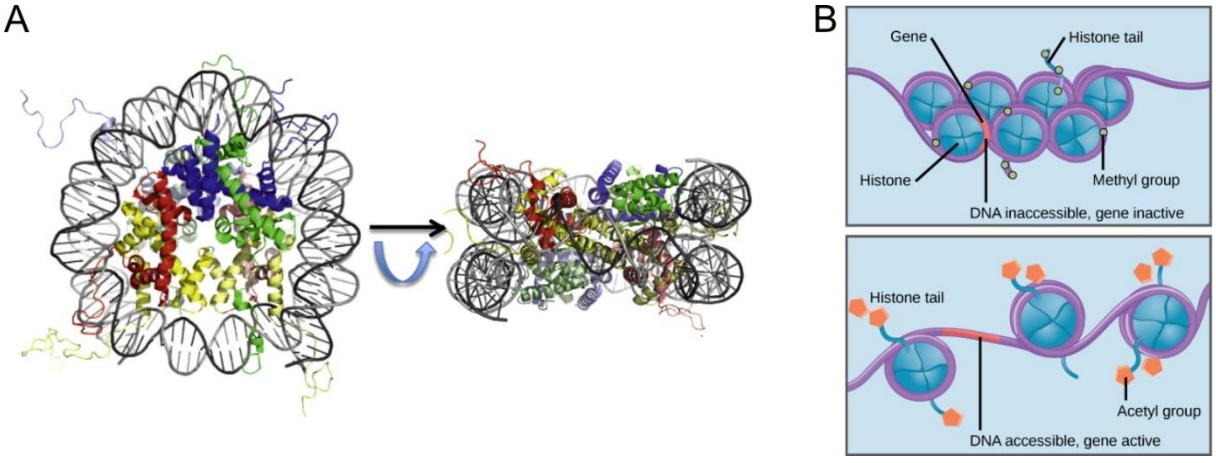



*Figure 2: The nucleosome structure. A) Nucleosomes are formed by 146 bp of DNA wrapped in 1.75 turns around a histone octamer with histone DNA contacts occurring roughly every 10 bp.*

*Histone octamers consist of two copies H3 (yellow) and H4 (red) forming a tetramer closer to the centre of the structure and two copies of H2A (green) and H2B (blue) forming two dimers. Figure from (81). B) Histones contain basic tails, which can be posttranslationally modified, which influences nucleosome spacing and therefore gene accessibility. Methylation of certain histone tails is most often associated with a more condensed chromatin state switching off nearby genes. Additionally, methylation of DNA strengthens electrostatic interactions with histones, leading to a higher occupancy. Acetylation of histone tails weakens histone-DNA contacts leading to a more open chromatin state allowing transcription of nearby genes. Figure from (82) (https://openstax.org/books/biology-2e/pages/16-3-eukaryotic-epigenetic-gene-regulation).*

Histones contain basic tails, protruding from the nucleosome core, and the post-translational modifications (PTMs) of these tails control chromatin folding and nucleosome self-association (83–85). Histone acetylation, regulated by histone acetyltransferases, weakens the binding to the negatively charged DNA and also reduces interactions between nucleosomes, leading to a more open chromatin formation called euchromatin with increased accessibility for other DNA-binding proteins (6, 83, 86, 87) **(Figure 2B)**. Therefore histone acetylation most often occurs at transcriptionally active genomic regions and leads to increased accessibility for TFs at transcriptional regulatory elements like promoters and enhancers and therefore to increased transcriptional activity for example at genes required for cellular differentiation (88–92). Additionally, histone acetylation leads to enhanced recruitment of chromatin remodelling complexes (CRCs) involved in the formation of more accessible chromatin (93, 94) **(Chapter I 1.1.3)**, which further increases accessibility for polymerases and TFs (94, 95).

Histone methylation can lead to both increased or decreased transcriptional activity dependent on the methylated amino acid residue. On the one hand trimethylation of lysine 4 in histone H3 results in interactions with parts of the transcription machinery and remodellers involved in creating an open chromatin state allowing gene expression (96–99). On the other hand di- and trimethylation of lysine 9 (H3K9me2/3) of histone 3 creates binding sites for proteins writing further repressive chromatin marks (100) and for proteins involved in DNA compaction and nucleosome bridging leading to a more condensed nucleosomal structure called heterochromatin (22, 101–105) therefore reducing DNA accessibility for TFs **(Figure 2B)**. This condensed chromatin state and the reduced accessibility are required for the gene silencing in certain genomic regions for example at transposable elements (106), as expression of these repetitive sequences can lead to their repositioning (107) causing mutations and genomic instability (108). Heterochromatin formation is also required at centromeric regions (109) enabling the binding of proteins required for sister chromatid cohesion (110) and preventing chromosome segregation errors (111). Furthermore, gene silencing is required at pericentric regions, where gene expression is associated with cellular stress and diseases like cancer and age-related diseases (112–115).

Besides the modification of histones, genome accessibility is also regulated by the appearance of non-allelic histone isoforms, which differ in the amino acid sequence of their basic tails and therefore in PTMs (116). Histone variants, including for example H2A.X and H2A.Z (117, 118), can be switched into and out of already formed nucleosomes via histone dimer exchange (119–121), allowing the dynamic regulation of chromatin accessibility. In mammals DNA damage induces phosphorylation of serine 139 in H2A.X (122), forming immobile chromatin domains around double-strand breaks (DSBs) (123), to which it recruits DNA repair proteins like Rad50 and Rad51 (124, 125). H2A.Z also functions in DNA repair, as it is exchanged onto nucleosomes at DSBs, increasing chromatin accessibility and enabling further PTMs and the binding of repair proteins (126). Furthermore, the incorporation of H2A.Z is required for the RNA Pol II recruitment and therefore for the activity of certain enhancers (127) and H2A.Z deposition at promoters can enhance the binding of both transcriptional activators and repressors (128). Additionally to the genomic positioning of nucleosomes, the presence of different histone variants and PTMs therefore further impacts DNA

accessibility and different processes (6) like transcriptional regulation (129, 130), DNA replication (131), DNA repair (132), and cellular differentiation (133).

### *1.1.3 Chromatin remodelling*

Nucleosome positioning influences multiple processes like transcription, during which the first nucleosome within the gene often causes Pol II pausing (134), and replication by regulating the accessibility of DNA for components of the replication machinery (135). The positioning and spacing of nucleosomes at certain genomic sites, like enhancers, promoters, replication origins as well as the exchange of histone variants are controlled by CRCs (37, 136). CRCs contain Snf2-type ATPase subunits, allowing them to perform their various functions by translocating along the DNA (37, 137, 138). Upon interactions with histones, they can weaken histone-DNA contacts (139) and can slide histones along the DNA (140), remove them from DNA (141) or exchange histone variants (119, 142). For these functions, they contain domains for nucleosome binding, recognition of histone variants, ATPase regulatory domains, and domains that allow them to interact with other chromatin regulatory factors or TFs (136).

They are involved in transcriptional regulation by accurately positioning highly conserved nucleosomal arrays downstream (+1,+2,+3…) and upstream (-1,-2,-3 …) of the TSS, controlling the formation of NFRs around promoters (5, 59, 143, 144). Nucleosome remodelling is required for the binding of TFs to their promoter binding site in both *yeast* and mammals (145–147). In higher eukaryotes, accurate nucleosome spacing is also required for epigenetic regulation as differences in DNA accessibility for TFs binding to enhancer elements can lead to changed gene expression patterns in different cell types (148). In *yeast*, Rap1 cooperates with the CRC RSC to remove nucleosomes from promoters creating NFRs (149–151), showing that TFs can also regulate nucleosome positioning. Accurate positioning of histones around TSSs leads to an increase in RNA polymerase binding (52), further displaying its requirement for transcriptional activation.

Apart from transcription nucleosomal occupancy also influences replication, as conserved spacing of nucleosomes is also found at replication origins (152), and an open chromatin state leads to efficient replication initiation (135). Furthermore, during DNA replication the CRCs Isw1 and Chd1 are required for the repositioning of nucleosomes behind the replication fork (153). Additionally, CRCs are required for the removal of nucleosomes from DSBs, allowing the loading of DNA repair proteins (154). Nucleosome architecture and the activity of CRCs, therefore, play important roles in multiple genomic processes, including transcription, DNA replication and DNA repair.

## 1.2 The roles of CTCF and cohesin in genome organisation

### *1.2.1 TAD formation*

During genome organisation, the chromatin fibre folds into loops, which are involved in the formation of TADs (13, 155). TADs are defined by more frequent 3D DNA-DNA interactions happening within a TAD domain compared to interactions occurring with neighbouring domains (16, 17, 156). TADs can therefore down- or upregulate expression by controlling interactions between regulatory DNA elements in 3D space (157). An example is interactions between promoters, start points of transcription initiation, and enhancers, which can be positioned far away from the promoter and regulate transcription by long-range DNA interactions, by recruiting additional TFs, CRCs and components of the transcription machinery (158–161). On the one hand, the folding of DNA into TADs and resulting decreased genomic distances of regulatory domains enables enhancer-promoter colocalisation and gene expression (16, 162–169). On the other hand, TAD boundaries can downregulate genes by preventing the interaction of promoters and enhancers not positioned within the same TAD (170–177). Loss of TAD boundaries can consequently lead to the activation of previously silent genes and is therefore associated with phenotypic changes

involved in a variety of diseases (178) like for example cancer in case of oncogene activation or defects in extremities formation in case of changed expression patterns of signalling factors involved in embryogenesis (172, 176, 179–181). Genome contact map data has shown that TADs are formed by the interaction of two proteins, CTCF and cohesin (18, 19, 40, 155), which will be further discussed below.

### *1.2.2 Cohesin entraps DNA in an ATP-dependent manner*

The structural maintenance of chromosome (SMC) complex cohesin (16, 156) is a tripartite ring-shaped protein complex formed by the ATPase head, coiled-coil and hinge containing subunits SMC1 and SMC3 and the α-kleisin subunit RAD21 (4, 182) **(Figure 3A)**. A HEAT-repeat subunit SA1 or SA2 binds cohesin by interactions with the RAD21 and SMC subunits (183, 184) and is required for targeting certain genomic sites like telomers, centromeres and enhancers (185–187). Cohesin gets recruited by another HEAT-repeat containing protein the cohesin loader complex Scc2/4 (NIPBL/MAU2 in *humans*) to centromeres (188), from where it translocates with an ATP-dependent mechanism to pericentromeric regions (189). Apart from Scc2/4 (190), topological cohesin loading (entrapment of DNA within the cohesin ring) requires functionally active SMC ATPase subunits (191). Whether topological loading is facilitated by hinge-opening (192) or if the RAD21-SMC3 interface functions as a gate (193, 194) is however still under debate.

Cohesin has been shown to be required for TAD formation as the number of loops and the amount of intra-TAD contacts drastically decrease upon degradation of the cohesin's RAD21 subunit (18, 155) **(Figure 3B)**. Cohesin's role in TAD formation requires an ATP-dependent process called loop extrusion (13), which has been demonstrated recently *in vitro* (14, 15) and depends on the cohesin loader. During loop this process, cohesin captures small DNA loops and gradually enlarges them moving along the DNA (14, 15). This DNA movement is different to well-studied DNA translocases, like polymerases or helicases (195, 196), which take only steps of one base pair at a time. Cohesin can move in steps of hundreds of bps (12) and is able to bypass obstacles attached to the DNA (197, 198). This indicates that cohesin does not move directly along the DNA strand, but instead overcomes large distances and DNA-bound obstacles by binding a second distant DNA position.

As cohesin's function is essential for genome architecture and TAD formation (18, 155) loop extrusion and DNA unbinding are regulated by the interaction with additional binding partners similar to cohesin loading. NIBPL can be replaced by the HEAT-repeat containing PDS5, which in complex with WAPL removes cohesin from DNA (18, 199) by creating an exit gate via dissociation of the N-terminal part of RAD21 from SMC3 (193). Depletion of WAPL therefore allows cohesin to be longer attached to and consequently form larger loops on genomic DNA (200). CTCF creates anchor points for loops created by cohesin, by a direction-dependent binding to its CTCF-binding sites (CBSs), explained in further detail below (12, 13, 18, 19, 40, 201).

### *1.2.3 CTCF targets specific sequences*

The transcriptional regulator CTCF has originally been identified as an insulator, blocking promoter-enhancer interactions for example at the *β-globulin* locus (3) and the *Igf2/H19* locus (202, 203) by binding to specific target sequences positioned between promoter and enhancer. CBSs contain a core and an upstream motif as well as less sequence-specific downstream DNA (204, 205). CTCF engages these binding sites via its eleven zinc fingers (ZFs) (30, 206), which are flanked by two unstructured termini (29) **(Figure 3C)**.

The ZFs contain a C2H2 motif, found in many TFs targeting specific sequence motifs (207). Domains from this group contain a β-hairpin (consisting of two antiparallel β-strands) followed by an α-helix in a left-handed ββα-fold, with a Zn atom coordinated between two cysteine and two histidine residues (30, 207–209). The α-helix side chains form hydrogen bonds with the major groove of the DNA (207, 208), as is the case for CTCF ZF 3-7, which follow the twist of the DNA and interact with 2 or 3 bases of the CTCF core motif each (30, 206).

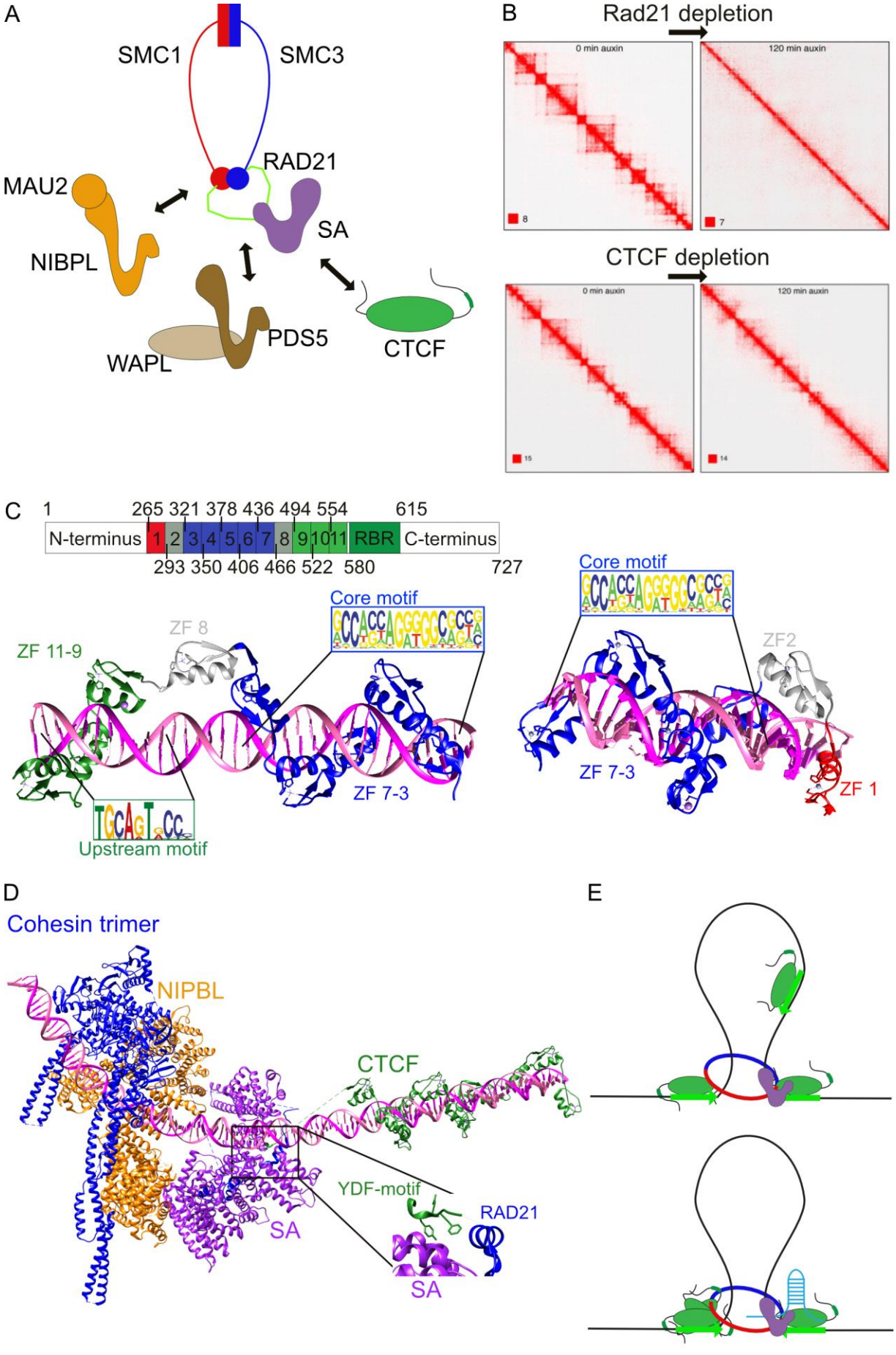

A
SMC1
SMC3
RAD21
MAU2
SA
NIBPL
PDS5
WAPL
CTCF
B
Rad21 depletion
0 min auxin
120 min auxin
CTCF depletion
0 min auxin
120 min auxin
C
1
265 321 378 436 494 554
615
N-terminus
1 2 3 4 5 6 7 8 9 10 11
RBR
C-terminus
293 350 406 466 522 580
727
Core motif
ZF 8
ZF 11-9
ZF 7-3
Upstream motif
Core motif
ZF2
ZF 7-3
ZF 1
D
Cohesin trimer
NIPBL
CTCF
YDF-motif
RAD21
SA
SA
E

*Figure 3: CTCF and cohesin shape TADs. A) Cohesin subunits SMC1, SMC3 and RAD21 form a ring structure, which additionally associates with the HEAT-repeat protein SA. Interaction with NIBPL/MAU2 enables cohesin DNA loading while the interaction with PDS5-WAPL can lead to cohesin unbinding. CTCF can interact with cohesin via SA and Rad21. B) Genome contact maps before and after 120 min of auxin-induced degradation of RAD21 for human chromosome 12 (top) or CTCF for human chromosome 4 (bottom). Figure from (18). C) Top: CTCF consists of eleven ZFs flanked by two unstructured termini. The C-terminus contains an RNA-binding region (RBR). Bottom: X-ray structures of CTCF ZFs 3-11 and CTCF ZFs 1-7 binding to DNA (PDB: 8SSQ and 8SSS (30)), with ZF 1 (red) binding downstream DNA, ZFs 3-7 (blue) binding the major groove of the CTCF core motif and ZFs 9-11 (green) binding an upstream motif. Upstream motif figure from (210). Core motif figure from (205). D) Cryo-EM structure of the human cohesin-CTCF-DNA complex (PDB: 7W1M (201)). The DNA molecule is bound by the cohesin trimer and NIPBL, both interacting with the HEAT-repeat protein SA1, which itself loosely binds the DNA via three positively charged regions. CTCF binds the CBS on the other end of the DNA with its eleven ZFs in a direction-dependent manner, causing CTCF's N-terminus to point towards cohesin interacting with the SA1 and RAD21 subunits via its YDF motif (201). E) Top: Cohesin extrudes DNA into loops until it encounters a CTCF molecule but only if it reaches the N-terminus of CTCF first (convergence rule). Bottom: Apart from a direct biochemical interaction, CTCF might perform its role in TAD formation by sterical blocking via oligomerisation (left) or RNA binding (right).*

ZF8 interacts with the phosphate backbone for 7-8 bp, acting as a spacer and allowing ZFs 9-11 to form hydrogen bonds with the major groove of the upstream motif via α-helix side-chains, with only 1 or 2 bases each (30, 206). ZF1 interacts with the major groove at a downstream triplet, while ZF2 also acts as a spacer and does not form any bp-specific interactions (30).

Across the genome, CBSs can often be found at TAD boundaries (16, 40), which leads to the assumption that CTCF has an important role in TAD formation. In line, depletion of CTCF decreases TAD boundary strength as in genome contact maps corner dots mostly disappear (18, 19). Additionally, contacts are not as clearly restricted to specific domains anymore leading to a less defined square structure **(Figure 3B)** (18, 19), showing that while cohesin's main function is in decreasing genomic distances within TADs, CTCF's main function is to act as a boundary factor. Deletion of CBSs consequently leads to boundary loss and the creation of new promoter-enhancer contacts with the above-mentioned effects on phenotypic changes and disease-related gene expression (171–173, 175, 176, 179, 181).

### *1.2.4 Loop anchoring*

It is still under debate how CTCF acts as a boundary factor and constrains loop extrusion, as it could depend on a direct biochemical interaction between CTCF and cohesin or on CTCF creating a physical barrier preventing the cohesin ring from moving past it. CTCF's high sequence specificity allows it to bind in a direction-dependent manner (205, 206) and most CBSs at TAD boundaries are oriented convergently with CTCF's N-terminus pointing towards the inside of the TAD (40, 211) **(Figure 3D,E),** suggesting that a biochemical interaction might depend on CTCF's N-terminus. Consequently, a direct interaction between CTCF's N-terminus with cohesin's RAD21 and SA subunit has been observed and mutations in CTCF's N-terminus impede TAD formation (201, 212, 213) **(Figure 3D)**, further showing that CTCF's N-terminus is involved in creating anchor points for cohesin loop extrusion. This SA-binding site is shared by CTCF and WAPL (212, 214), and CTCF can facilitate the formation of long-lived cohesin loops by preventing WAPL-mediated cohesin unbinding at CBSs (215), which proposes a mutually exclusive binding of these proteins to cohesin and might explain how CTCF stably anchors cohesin to its CBSs. However *in vivo* SA was shown to directly interact with CTCF in the absence of other cohesin subunits (216) and another study showed SA to interact with CTCF's C-terminus independently of RAD21 (217), unlike the Rad21-dependent N-terminal interaction (201, 212).

Apart from a direct biochemical interaction between cohesin and CTCF (212), CTCF might form a physical barrier for loop extrusion, as such a physical barrier has for example been observed by multiple closely bound Rap1 molecules blocking loop extrusion of the ring-shaped condensin complex (218). CTCF's termini form an elongated conformation (29) and could protrude from the DNA-bound ZFs blocking loop extruding cohesin. The cohesin ring in a fully open conformation has a 35 nm diameter (219), however, cohesin has also been shown to form a collapsed state with a smaller diameter (197), in which it was able to move past small DNA-bound obstacles of 10.6 nm diameter but not past larger obstacles of 19.5 nm diameter or more (197). In both cases, the termini (N-terminus 10.2 nm diameter, C-terminus 7.4 nm diameter) of a single CTCF molecule are most likely too small to block cohesin. Another study showed that cohesin can incorporate even 200 nm particles into its formed loops (198), further excluding a physical barrier by a single CTCF. Additionally, single CTCFs were found to be imperfect barriers (12, 220), indicating that approaching cohesin complexes do not necessarily have to bind CTCF even if approaching it from the N-terminal site. CTCF oligomerisation **(Figure 3E)** observed *in vitro* (221, 222) and *in vivo* (223, 224) could increase barrier strength by forming large physical barriers preventing the cohesin ring from moving past them. CTCF oligomerisation has also been linked to RNA binding (225, 226) and recruiting RNA could be another mechanism for how CTCF blocks cohesin, possibly by creating a larger physical barrier binding long RNAs (227) or by creating a biochemical interaction with cohesin's RNA-binding SA subunit (216, 228) **(Figure 3E).** Stalling of loop extrusion at CBSs and therefore TAD insulation could therefore be a combination of multiple mechanisms including a direct biochemical interaction of cohesin with CTCF, which can stabilise cohesin on DNA by protecting it from WAPL (212, 214), sterical blocking by oligomerisation and the recruitment of additional cofactors like RNA.

In this thesis, potential mechanisms of CTCF anchor site formation are analysed, showing a direct interaction between CTCF's ZFs and SA as well as a CTCF oligomerisation-dependent mechanism for RNA capture (229). Additionally, it is shown that cohesin can form force-resistant bridges on singular DNAs in cis in an ATPase-dependent manner, revealing a potential mechanism for cohesin loop formation (230).

### *1.2.5 Cohesin's roles in sister chromatid cohesion and DNA repair*

Apart from its role in TAD formation, cohesin also has other important roles in cell division and DNA repair (20, 231–233). During S-phase, acetylation of cohesin leads to the replacement of NIBPL by Pds5 (234). Cohesin subsequently holds sister chromatids together inside its ring structure until the onset of anaphase (20), when the protein ring is cleaved by separase (235). To perform sister chromatid cohesion, cohesin must, therefore, be able to bridge two DNA molecules in trans. Thereby, either a single cohesin complex might entrap both DNA molecules inside its ring in the "embrace" or "ring" model (182) or two interacting complexes might entrap one sister chromatid in the "handcuff" model (236). In the first model, cohesin must be able to either allow passage of a second double-stranded DNA (dsDNA) into its ring or capture a single-stranded DNA (ssDNA) during DNA replication (237), while the second depends on cohesin-dimer formation via its Rad21 subunit (236).

Cohesin is enriched at DSBs, depending on its interaction with the cohesin loader, DNA repair proteins and the phosphorylation of histone H2A.X in *yeast* (232, 238). How exactly cohesin functions in DNA repair remains unknown, but by holding sister chromatids together and reducing DNA mobility, cohesin may reduce the risk of using non-sister chromatid DNA as donors during recombination (232, 239). Interestingly cohesin can still perform this function of keeping sister chromatids together for DNA repair during late mitosis even after the kleisin Scc1 is cleaved for sister chromatid segregation, which most likely requires loading of additional cohesin onto chromatin containing uncleaved Scc1 (240). Additionally, cohesin enables DNA repair by keeping the DNAs on both sides of the DSB in close contact by entrapping them within its ring (241).

This thesis contains mechanistic insights into how cohesin might hold two different DNAs at DSB sites (241) and sister chromatids together, required for directed DNA damage repair (239) and preventing early separation of sister chromatids during cell division (20). Cohesin is shown in **Chapter II 3.** to bridge two parallel DNAs in the presence of Scc2/4 and ATP. These tethers formed by cohesin can move along the two DNAs and resist high forces (230).

### *1.2.6 CTCF functions in epigenetic regulation*

While CTCF's genomic function in TAD boundary formation depends on targeting its CBS (13, 16, 18, 40), it is further influenced by additional DNA- and RNA-binding properties of CTCF as well as PTMs. CTCF genomic binding positions differ between cell types depending on DNA modifications such as DNA methylation (242), which causes changes in CTCF binding during cellular differentiation of *mouse* embryonic stem cells and can lead to CTCF boundary loss and increased gene expression in differentiated cells (7). Two of the cytosines within the CTCF core motif are part of a CpG or CpA dinucleotide, meaning that the 5-carbon atom of the cytosine is often methylated *in vivo* (243–245). Methylation of one of the two sites leads to a significant reduction in CTCF binding affinity (244), which can lead to loss of CTCF-mediated boundaries and increased oncogene expression in some cancer cell lines (180, 246). Additionally, it reveals a function of CTCF in genomic imprinting for a CBS positioned between the H19 enhancer and the *Igf2* gene (202). In the maternal allele, CTCF binds the unmethylated CBS, blocking the promoter-enhancer contact inactivating the *Igf2* gene and leading instead to activation of the more close-by *H19* gene. In the methylated paternal gene, CTCF binding is inhibited, leading to *Igf2* gene activation by the H19 enhancer (202). Furthermore, DNA methylation plays a role in alternative splicing. CTCF can cause Pol II pausing at the CD45 exon 5, leading to exon inclusion, which is inhibited by DNA methylation preventing CTCF from binding to its CBS (247). CBS methylation can therefore lead to changes in gene expression in different cell types by influencing CTCF boundary function as well as the generation of different protein isoforms by influencing RNA splicing.

In addition to DNA modifications, CTCF is also regulated by different PTMs. Phosphorylation of CTCF's C-terminal domain does not influence DNA binding *in vitro* but leads to changes in gene expression, suggesting that it might be involved in interactions with other proteins (248). Stress-induced phosphorylation of CTCF's ZFs on the other hand impedes DNA binding to some of its CBSs preventing loop-formation and activation of genes involved in cell growth (169). Another common CTCF PTM poly(ADP-ribosyl)ation of CTCF's N-terminus explains the presence of two molecular weight variants 130 kilodalton (kDa) CTCF and 180 kDa CTCF, the first containing no or only a few and the second multiple ADP-ribose residues (249, 250). CTCF's CBSs are bound by both variants, but there are also genomic binding sites without a CBS bound by only the CTCF 180 variant, suggesting that this variant can bind to and control gene expression at some genomic regions by binding a different motif or by being recruited by other proteins (250). This transcriptional regulation by different CTCF variants most likely includes genes involved in growth control as CTCF 180 is exclusively present in healthy *human* breast cells, which are cells associated with low proliferation, whereas tumour cells contain a large amount of the less ADP-ribose containing CTCF 130 (251, 252). CTCF and its PTMs therefore play important roles in regulating cellular signalling pathways and proliferation.

### *1.2.7 Most of CTCF's genomic functions are influenced by RNA binding*

Additionally to its DNA binding function, CTCF also interacts with RNA via its ZFs 1,10 and 11, an RNA-binding domain on its C-terminus and possibly also via central ZFs (112, 225, 226, 253, 254). Interaction with RNA is required for many of CTCF's roles in X-chromosome inactivation, gene regulation and genome architecture (227, 253).

The choice of the inactive X-chromosome appears to be random in placental mammals (255) but requires the interaction of the two X-chromosomes (256), mediated by CTCF binding to non-coding RNAs (ncRNAs) (227). During X-chromosome inactivation the long ncRNA Xist covers the X-

chromosome, leading to repression of transcription (257, 258). A positive regulator of X-chromosome inactivation the long non-coding RNA (lncRNA) Jpx can bind to CTCF and thereby cause CTCF unbinding from the the *Xist* promoter, leading to Xist expression and the initiation of X-chromosome inactivation (259).

Jpx RNA can therefore compete with DNA for CTCF binding and has been shown to have an impact on TAD positioning being able to cause CTCF unbinding from genomic DNA depending on CTCF's affinity for the bound DNA sequence (260). On the one hand interaction with RNA can therefore lead to the removal of CTCF boundaries, while on the other hand, it can also be required for boundary formation as the ncRNA SRA enables SA2-mediated cohesin-CTCF interaction at the *Igf2/H19* locus required for CTCF's function in preventing aberrant gene expression at this locus (261). Additionally to boundary formation, interactions with RNA are required for CTCF-mediated enhancer-promoter looping and gene activation, for example via R-loops formed by the ncRNA HOTTIP (166). RNA requirement for DNA loop and TAD formation is further displayed as the expression of ncRNAs close to CTCF boundaries leads to increased CTCF binding and increased gene expression within respective TADs while the ncRNA expression also leads to reduced interactions across CTCF boundaries (168).

Genome contact maps of RNA-binding deficient CTCF mutants display a loss of TAD boundaries resulting in changed gene expression due to both newly formed genomic contacts across CTCF boundaries as well as lost promoter-enhancer loops (226, 253). One of the studies showed that not all TADs are dependent on CTCF's RNA-binding function (226), suggesting different mechanisms for RNA-dependent and RNA-independent boundary formation at different TADs. CTCF has been shown to form large clusters with RNA *in vitro* (225) and RNA-binding deficiency reduces CTCF clustering *in vivo* (226), suggesting that RNA-dependent TAD formation involves CTCF oligomerisation.

This shows that interactions with RNA can both weaken CTCF's interaction with genomic DNA as well as enhance CTCF recruitment to certain sites and increase TAD boundary strength, depending most likely on CTCF's DNA affinity at the given interaction site and CTCF oligomerisation (168, 226, 260). In **Chapter II 4.** I show that CTCF can oligomerise in the absence of RNA but its interaction with RNA depends on CTCF oligomerisation, as monomeric CBS-bound CTCF was unable to perform secondary capture of RNA. Additionally, I show that CTCF binds to RNA transcripts on DNA curtains, however with a lower affinity compared to its CBS, suggesting that *in vivo* it might not only be recruited to CBSs but also to sites of active transcription.

### *1.2.8 CTCF- and cohesin-related diseases*

Due to their important roles in genome architecture and transcriptional regulation, CTCF and cohesin play a role in multiple diseases. Mutations in cohesin or its associated proteins like the cohesin loader cause developmental syndromes called cohesinopathies, like the Cornelia de Lange Syndrome and Roberts Syndrome, associated with intellectual disabilities, growth defects leading to changes in stature and facial structures as well as defects in extremities formation (231, 262, 263). The Cornelia de Lange Syndrome is most often caused by dysregulated gene expression including for example genes required for nerve cell function and intellectual development or signalling pathways in embryonic development and proliferation and is usually associated with NIBPL mutations (263–265).

SA1 and SA2 containing cohesins display different genomic binding positions, with SA2- cohesin more often bound at sites of active transcription (186, 187) and cohesin's SA subunit is required for interaction with CTCF at TAD boundaries (212), showing a critical role of SA2 in regulating gene expression. Consequently, loss of or deficient expression of SA2 and subsequent changes in enhancer-promoter contacts leading to increased oncogene expression are involved in the formation of different forms of cancer (266, 267).

Mutations in CBSs lead to changes in TAD boundary positions, promoter-enhancer interactions and therefore in gene expression (172, 176, 178, 179, 181, 268). CBS mutation-associated expression changes of transcriptional regulators, receptors and signalling factors involved in cellular differentiation and proliferation during embryonic development lead to defects in extremities formation (172, 268). Changes in oncogene expression lead to different types of cancer (179, 269) and can be used to identify DNA positions that are frequently mutated in cancer patients by machine learning (270).

Loss-of-function mutations of CTCF are rare as they probably cause low survivability (271, 272). Due to dysregulations in interactions between enhancers and gene promoters of a wide range of genes, CTCF mutations can cause a variety of disorders (272) like microcephaly, intellectual disability and growth retardation (272–274) as well as different forms of cancer (275, 276).

A better understanding of how CTCF and cohesin perform their functions inside the nucleus regulating genome architecture and transcription is therefore critical for the treatment of various diseases.

## 1.3 Higher-order compartments and chromosome territories

### *1.3.1 A- and B-compartments*

Additionally to TADs, genome organisation is shaped by two compartments containing transcriptionally active and inactive chromatin (21, 277). The compartmentalisation of DNA was first detected almost a hundred years ago, when microscopy revealed two different structures, euchromatin and heterochromatin, with the second being more compacted, visible as dark condensed spots of DNA (278).

Transcriptionally inactive regions, marked by repressive histone modifications such as H3K9me2/3 and H3K27me3 (279, 280) and hypoacetylation (281), causing less accessibility for other proteins (282) and a more condensed structure (283) are defined as heterochromatin (280, 284–286). Heterochromatin forms self-interacting domains at the nucleolus and nuclear lamina called nucleolus-associated domains (NADs) and lamina-associated domains (LADs), which are rich in repetitive sequences and repressive histone marks and contain fewer and less active genes than euchromatin regions (41–43, 163, 287, 288).

Gene-rich euchromatin is more often positioned at the nuclear interior (289–293) and associated with active genes marked by histone acetylation (294–298). Acetylation prevents DNA compaction (83, 299, 300), allows increased TF and CRC binding (92, 301, 302) and leads to enhanced transcription (299, 303, 304).

This spatial separation of euchromatin (nuclear interior) and heterochromatin (nuclear lamina and nucleoli) (42, 291) can be observed by the formation of two compartments (A/B-compartments containing mostly euchromatin/heterochromatin respectively) within individual chromosomes in genome contact maps (21, 277). These compartments form less frequent interactions with neighbouring regions (regarding the position along the DNA strand) of the opposite epigenetic type and long-range interactions with regions of the same epigenetic type (21, 277) **(Figure 1C).**

### *1.3.2 Chromosome territories*

At the largest genome organisation scale, chromosomes fold into distinct territories, containing condensed chromosomal DNA separated by interchromatin space (305). These territories of individual chromosomes to a lesser extent also form interchromosomal interactions between certain chromosome pairs (306). Similar to compartments, the distribution of chromosome territories inside the nucleus is linked to transcriptional regulation, as chromosomes containing a

high amount of genes are found in the inner parts of the nucleus, while territories with poor gene density are more often associated with the nuclear lamina (37, 38, 289, 307–310). In the genome contact map, most signals are restricted to squares on the matrix, displaying the more frequent interactions within individual chromosomes and less frequent interchromosomal interactions **(Figure 1B)** (39, 311).

### *1.3.3 Compartments are shaped by phase separation*

In contrast to TAD formation, which as explained above is caused by CTCF boundaries and by cohesin-mediated loop extrusion, the formation of higher-order chromatin compartments seems to be mainly driven by phase separation (22, 23, 32, 312, 313) and independent of cohesin, cohesin cofactors and CTCF (18, 155, 163, 314). Proteins inside the nucleus can form dense and compact networks, segregated from the surrounding liquid, by forming multiple molecular interactions, including oligomerisation or recruitment of other proteins, RNA, nucleosomes and DNA (22, 23, 25, 27, 28, 315–317). In order for phase separation and therefore the formation of membrane-less compartments to occur the energetic gain from these molecular interactions must be larger than the loss of entropy from creating different phases (28, 317).

Phase-separated droplets form frequent regulatory interactions within them, while also leading to less frequent interactions with molecules outside of them, causing the separation of different chromosomal compartments (22, 27, 28, 312, 318). On the one hand, this promotes frequent genomic interactions of components of the transcription machinery in regions of euchromatin facilitating gene expression in these regions (25, 313, 319) **(Figure 4)**. On the other hand reducing the access of euchromatin-associated factors by DNA compaction, condensate formation and increased interactions between nucleosomes can play a role in gene silencing in heterochromatin regions (22, 23, 105, 320–322). Silencing of heterochromatin regions is important for genome stability as expression of transposable elements can cause their repositioning and thereby induce genomic mutations (106–108), while aberrant expression in centromeres and telomeres can cause defects during cell division (323–325). Changes in transcription of these regions can lead to different diseases like cancer (101, 326) and age-related diseases (112, 327).

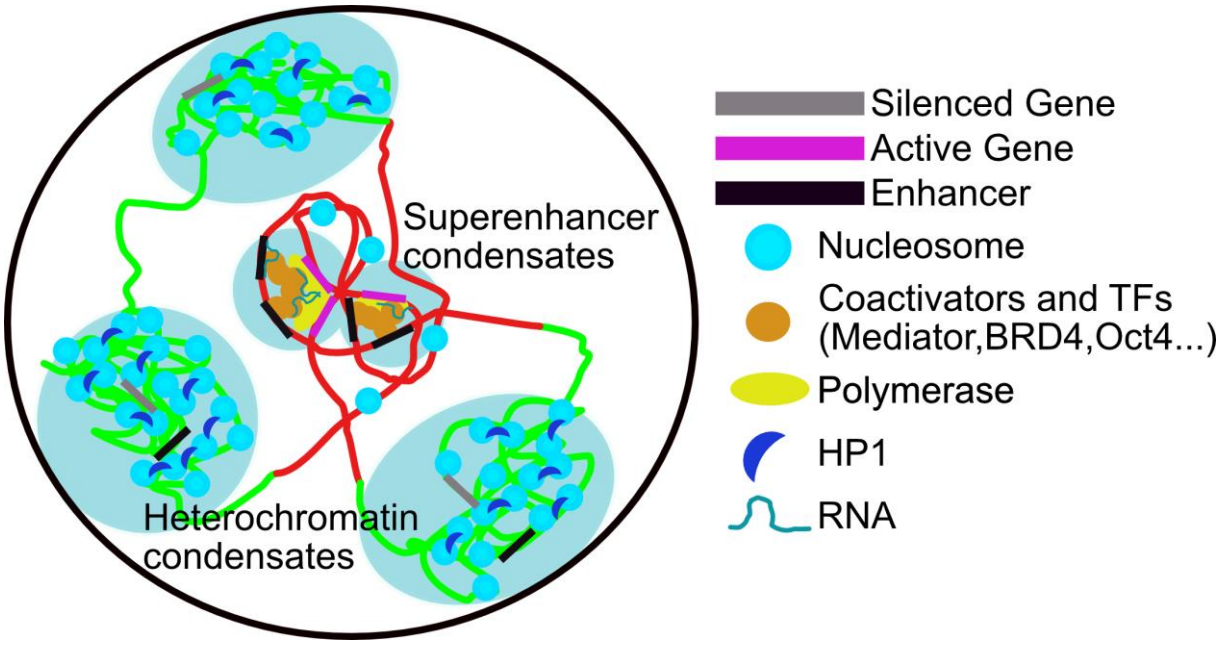



*Figure 4: The role of phase separation in genome compartmentalisation. Different epigenetic regions along the DNA strand (euchromatin: red, heterochromatin: green), are separated from each other presumably by the formation of phase-separated condensates leading to increased interactions within these genomic regions and reduced interactions with the surrounding solvent.*

*Heterochromatin is dense in nucleosomes, bridged by HP1 and displays strong DNA compaction. Euchromatin is in a more open conformation, which enables gene expression by condensate formation mediated by transcriptional coactivators and RNA Pol II. Figure based on (27).*

Important for heterochromatin-associated condensate formation is HP1, which forms phase-separated droplets and causes DNA compaction *in vitro* by bridging DNA via its positively charged and disordered hinge and forming dimer-dimer contacts via phosphorylation-dependent interactions between its hinge and N-terminus (22, 23, 328–330). HP1 colocalises with heterochromatin regions via interactions with trimethylated histones H3 (H3K9me3) (101–103), a common mark for repressive chromatin, suggesting that its phase-separation and DNA compaction properties are required for the maintenance of condensed heterochromatin *in vivo* and the reduced accessibility of heterochromatin (322). The isoform HP1β forms condensed complexes selectively with H3K9me3 containing chromatin *in vitro* (105), which might lead to further compaction of heterochromatin regions.

Another example is the Polycomb-repressive complex 1, which can form droplets *in vitro* (331) and its condensates colocalise with the repressive H3K27me3 histone PTM and lead to the formation of further repressive chromatin marks, suggesting that it is involved in the spreading of heterochromatin (332). It can also lead to the formation of large clusters by forming bridges to other Polycomb-repressive complex 1 containing regions (333), which could therefore contribute to the genome-wide compartmentalisation into A- and B-compartments.

MeCP2 forms dense clusters with nucleosome-containing DNA *in vitro* (334) and forms condensates in the presence of DNA, which preferentially incorporate proteins associated with heterochromatin phase separation over proteins associated with euchromatin phase separation (321), suggesting that it might prevent the access of transcriptional activators to regions of heterochromatin.

Heterochromatin is, therefore, maintained by interactions between histone tails and proteins involved in phase separation processes leading to the formation of dense compacted networks reducing DNA accessibility and excluding euchromatin-associated factors (22, 23, 27, 277, 321, 335).

## 1.4 Interplay of transcription and genome architecture

### *1.4.1 Impact of transcription on compartments*

Phase separation is not limited to repressed B-compartments and can be shaped by transcriptional processes (27, 28, 32, 312). Condensates often contain RNA-binding proteins (RBPs) (336), which form hydrophobic interactions with themselves, other RBPs and transcription-associated factors and electrostatic interactions with transcription-produced RNA (28). One example is the transcriptional regulator FUS, which contains an RNA-binding domain required together with RNA for phase separation and a prion-like domain influencing droplet size (315). FUS interacts with the C-terminal domain of RNA Pol II (337) and incorporates it into its formed droplets *in vitro* (338). Since condensate-dependent phosphorylation of RNA Pol II's C-terminal domain is required for forming an actively transcribing elongation complex (EC) (339–341), this suggests a role of FUS condensates in transcription activation.

A type of ncRNA produced at enhancers (eRNA) can increase promoter-enhancer contacts and stimulate expression of nearby genes (342–344), which might require additional transcriptional cofactors as the transcriptional regulator BRD4 is recruited to enhancers by RNA leading to gene activation (345). BRD4 and MED1, a subunit of the large transcriptional coactivator complex

Mediator, contain large IDRs and form condensates *in vitro*, with MED1 also being able to form dense Pol II and BRD4 enriched condensates in nuclear extracts (24, 319). These condensates are also relevant for the function of these proteins *in vivo* as the Mediator complex and Pol II enrich in large stable protein clusters in living cells (25). They are involved in the formation of super-enhancers (313) **(Figure 4)**, characterised by high cooperativity between multiple enhancers, increased gene expression and transcriptional bursting and often associated with cell-type specific genes (313, 346, 347). RNA Pol II can form condensates via its intrinsically disordered C-terminal domain (CTD) together with proteins causing CTD phosphorylation and enabling promoter release (339, 340) and the phosphorylated CTD to form condensates with proteins involved in RNA splicing (348). The TF Oct4 only forms droplets in the presence of MED1 *in vitro* and the TFs Oct4 and GCN4 colocalise with Mediator condensates *in vivo,* suggesting that these super-enhancer condensates recruit additional TFs for gene activation (316). Colocalisation of multiple enhancers, TFs, coactivators and polymerases at promoters can therefore cause the quick and strong activation of genes (burst expression) (313), which depends on close distances between the gene and the formed condensate (349). Cell-type specific gene expression in super-enhancers is mediated by TFs and transcriptional regulators involved in various cellular signalling pathways, which are recruited to Mediator condensates via their IDRs *in vitro* and colocalise with super-enhancer regions *in vivo* (350). Euchromatin is therefore organised by multiple contacts between factors associated with the transcription machinery, its produced RNA and their formed condensates (32, 163, 312, 313, 351).

CTCF can directly interact with Pol II (352), and CTCF oligomers are required for the formation of RNA Pol II transcriptional condensates (353). Super-enhancers controlling expression of cell-type specific genes in mammalian cells are frequently positioned between CBSs (354) and super-enhancer regions are enriched in both CTCF and cohesin (355) which raises the question, if CTCF's RNA-binding domains (226, 253) or IDRs (29) are directly involved in the formation of super-enhancer condensates. In this thesis, I analyse CTCF's ability to oligomerise and form droplets *in vitro*.

### *1.4.2 Impact of transcription on TADs*

Additionally to compartment formation, transcription can also influence TAD formation (163). Organisms which do not express CTCF (e.g. *C. elegans* and different species of *yeast* and *bacteria*) still form self-interacting domains (163, 356–361), and their boundaries are often shaped by replication or transcription processes (44, 356–358). In organisms that contain CTCF like *Drosophila*, most TADs do not depend on the presence of convergently orientated CBSs and are unaffected by CTCF loss (44, 362) and some TAD boundaries remain after CTCF loss in *mice* (19), while in both cases many TAD boundaries are found at promoters of highly active genes (16, 44, 163, 363, 364), suggesting that transcriptional processes or produced RNA might take over the role of CTCF in TAD boundary formation. In *humans,* a combined role of transcription and CTCF might shape TAD formation as transcription of ncRNAs that can interact with CTCF leads to less frequent interactions across CTCF boundaries as well as to more frequent enhancer-promoter interactions within TADs and therefore to the activation of genes within respective TADs (168).

At the same time, transcription and transcription-produced RNAs can also cause CTCF unbinding and thereby indirectly determine TAD positioning by removing CTCF from some boundaries (259, 260, 365). Additionally, simulation data combined with HiC experiments suggest that transcribing RNA polymerases might take over the role of CTCF in the positioning of TADs by acting as boundaries slowing or stopping loop-extruding cohesin and even repositioning it during transcription (366). This is consistent with previous results which show that cohesin positioning depends on active transcription as it might be pushed by polymerases after initial loading in both *yeast* and mammals (163, 367–369). Moreover, transcription can influence DNA accessibility, as removal of cohesin from CBSs by transcribing polymerases can lead to a more open chromatin

formation (370). This means that transcription and produced RNAs influence cohesin positioning (368), CBS selection (259, 260) and TAD boundary strength (168) and therefore regulate CTCF and cohesin-mediated genome architecture.

In this thesis, a single-molecule transcription assay was developed to directly study the impact of a transcribing RNA polymerase on architectural proteins like CTCF or SA. Additionally, the recruitment of these proteins by transcription-produced RNA was observed as well as eviction from DNA by RNA in the case of SA, displaying how transcription and transcription-produced RNAs can influence the localisation of architectural proteins.

### *1.4.3 Polymerase pausing and protein roadblocks*

While transcription plays a role in shaping TAD/LAD boundaries and chromatin compartments, nucleosomes stay attached to chromatin during transcriptional processes (371–373), which raises the question of what happens when transcribing polymerases encounter roadblocks like nucleosomes (374) but also other roadblocks like TFs, genome architecture proteins or stalled polymerases.

RNA release frequently occurs during the early stages of transcription before the transition from the RNA polymerase initiation complex (IC) to the efficiently transcribing EC in a process called abortive cycling (375–379). It can be caused by the reannealing of DNA at the downstream end of the transcription bubble (380), reducing the length and therefore the stability of the DNA:RNA hybrid or by a mechanism in which the polymerase enlarges the transcription bubble by pulling downstream DNA past its active centre without moving on DNA causing instability of the IC (381) **(Figure 5A)**. Additionally, it has been shown that abortive cycling can also include dissociation of the polymerase (382). It has been observed in eukaryotes (376), bacteria (378) and viruses (380) and depends on promoter sequence as purine-rich segments in the non-template strand stabilise the IC and reduce abortive cycling (378). Transition into a more stable EC and promoter escape require multiple conformational changes including reannealing of the upstream part of the transcription bubble causing a release from promoter-bound TFs in eukaryotes (376).

After promoter escape eukaryotic polymerases can be caused to pause by a promoter-proximal CTCF site (383), and CTCF plays a role in the recruitment of two cofactors NELF and DSIF known to cause an early pause during transcriptional elongation required for RNA processing (384). NELF causes the polymerase to pause by binding to the polymerase funnel (385) and DSIF halts the EC by tightly binding the exiting RNA and DNA (386) **(Figure 5B)**. CTCF also causes the binding of P-TEFb, which induces a restart by phosphorylation-induced pause factor release (383, 387–389). Interestingly NELF cannot only downregulate transcription by causing Pol II pausing but also lead to enhanced gene expression as NELF-mediated Pol II pausing is associated with decreased nucleosome formation at promoter regions and an increased presence of nucleosome modifications associated with open chromatin (390).

Promoter proximal pausing is also influenced by nucleosome positioning as insertion of a promoter-proximal artificial 601 site led to increased pausing and NELF recruitment (391) and this pause factor recruitment might explain why the +1 nucleosome causes more frequent pausing than other nucleosomes within the gene (134). As nucleosomes within gene bodies are frequent obstacles for polymerases moving along the DNA track RNA polymerases have developed a mechanism to overcome them called template looping model (374) **(Figure 5C)**, in which the RNA polymerase destabilises some DNA-histone dimer interactions while pausing (392), which can cause the loss of an H2A/H2B dimer in some cases (393), followed by the formation of a DNA loop, which enables the histones to be repositioned past the polymerase from downstream to upstream DNA and new nucleosome assembly behind the polymerase (374, 393–396). While nucleosomes remain bound to DNA this leads to their repositioning (393, 394, 396–398). Accurate nucleosome spacing could then be reestablished by CRCs (59, 399). Polymerase encounters with nucleosomes are

additionally regulated by CRCs as FACT can reduce the pause timing, which requires an interaction with the H2A/H2B dimer, suggesting that FACT weakens dimer-DNA contacts (374, 400) and multiple CRCs are involved in accurate positioning of nucleosomes around the TSS required for efficient transcriptional elongation (59, 143, 144, 401).

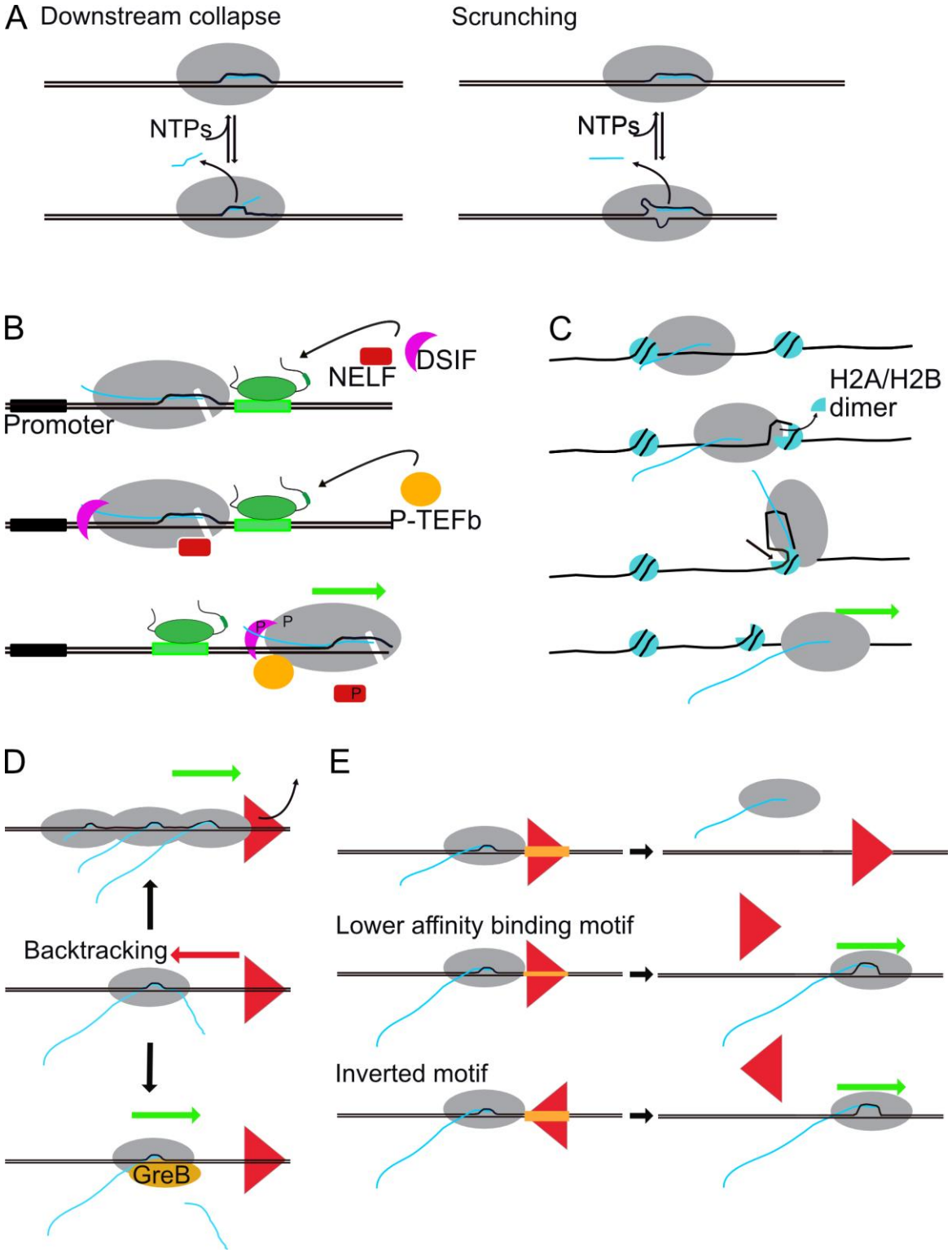



*Figure 5: Different pausing and unbinding events occurring during transcription. A) During initiation polymerases frequently undergo cycles of releasing short RNAs and restarting transcription. This can be caused by reannealing of the downstream end of the transcription bubble weakening DNA:RNA interaction. Alternatively, it can be caused by polymerases pulling in downstream DNA leading to a destabilisation of the transcription bubble. Figure based on (380, 381). B) CTCF can*

*regulate promoter-proximal pausing by recruiting NELF (red) binding to the polymerase funnel and DSIF (pink) binding exiting DNA and RNA causing pausing. CTCF-mediated P-TEFb recruitment leads to phosphorylation of the polymerase and pause factors causing unbinding of NELF and turning DSIF into an elongation factor. Figure based on (383). C) Polymerases frequently have to bypass nucleosomes during transcription requiring weakening histone-DNA contacts, which can potentially include H2A/H2B dimer loss, DNA looping and nucleosome reassembly on upstream DNA. Figure based on (374). D) Collisions with other proteins often cause backtracking of polymerases. This can be overcome by multiple polymerases working together pushing the backtracked polymerase forward or by additional cofactors stimulating RNA transcript cleavage by the polymerase and thereby causing realignment of RNA to the active site. Figure based on (402, 403). E) Polymerases are frequently caused to terminate upon collision with a roadblock. However, the lack of a strong binding motif or inversion of the motif has been shown to enable the polymerase to move past roadblocks (404).*

RNA polymerase pausing induced by a collision with nucleosomes or with other DNA-bound proteins (134, 405, 406) often results in backtracking with the 3' end of the RNA being no longer properly aligned to the template strand and the polymerase moving along the DNA strand without nucleotide addition (407) **(Figure 5D)**. Polymerases require a mechanism to realign the RNA to the active site (404), which requires cofactors like GreA and GreB in bacteria (403, 405, 408), which stimulate the endonucleolytic RNA transcript cleavage of the misaligned 3' end by the RNA polymerase with subsequent transcription restart upon polymerase movement by one base pair leading to the generation of a new active site (403). Alternatively, paused or backtracked polymerases can be assisted by trailing elongating polymerases pushing leading polymerases after roadblock collision and enabling them to move past the roadblock (402, 409).

Polymerase pausing is therefore regulated by nucleosome positioning (391) and transcription-associated factors (385), but can also be regulated by genome architecture proteins. Cohesin can influence promoter-proximal pausing by causing a reduced super elongation complex formation at promoters, which leads to impaired release from NELF-mediated pausing (410). Analysing genomic positioning of a transcribing polymerase at different time points showed that RNA Pol II often colocalises with and therefore presumably stops at CTCF- and RAD21-bound sites (411), in contrast to results suggesting the displacement of cohesin and in some cases also CTCF by transcribing polymerases (370). DNA-bound CTCF can cause Pol II pausing at a CBS, which leads to exon inclusion, while methylation of the CBS inhibits CTCF binding and consequently leads to exon exclusion (247). This shows that while transcription can shape genome architecture by influencing the genomic positioning of CTCF and cohesin **(Chapter I 1.4.2)** genome architecture proteins also directly influence the movement of transcribing polymerases.

Transcription termination includes slower transcription at the poly(A) recognition site, followed by cleavage of the RNA 5' end and degradation of RNA, causing a release of the less stably bound RNA transcript (412–416). DNA-bound roadblocks can however also lead to transcription termination (404), which can have important regulatory functions, as the TF Cbf1 acts as a roadblock binding to its target DNA at centromeres and preventing readthrough transcription into centromeres, while increased transcription would impair kinetochore formation (417, 418). Another example is the transcriptional activator Reb1p binding NFRs around TSS and preventing transcriptional readthrough from the previous gene via blockage of RNA Pol II and causing subsequent degradation of both RNA and polymerase (406). Roadblocking can depend on DNA sequence as in bacteria, it has been shown that roadblock-binding to low-affinity sites enables the polymerase to cause unbinding of the roadblock (419) **(Figure 5E)**. It can also depend on the orientation of sequence motifs (404), with the transcriptional regulator Rap1p reducing expression of a gene containing its binding site with the effect however being lost upon motif inversion, which seems to allow the polymerase to push off the roadblock (420).

Polymerases therefore frequently pause or are terminated when encountering roadblocks, which can have important regulatory functions for example preventing exon exclusion, protecting centromeres or preventing transcriptional readthrough, but also have mechanisms to bypass these obstacles for example by looping mechanisms or by cooperativity of multiple polymerases (247, 374, 402, 404, 409, 417). To study the influence of roadblocks on transcription, collisions between multiple polymerases were analysed in this thesis as well as collisions with architectural proteins like CTCF or SA (229), with different CBS orientations.

## 1.5 A single-molecule approach to study the different layers of genome architecture

To achieve a better understanding of how CTCF, cohesin, nucleosomes and transcribing polymerases shape the chromatin landscape and regulate gene expression, a more detailed understanding of how they individually interact with DNA is required **(Figure 6)**. Here, a method called DNA curtains can be applied to visualise hundreds of DNA molecules at the same time and study their interaction with individual proteins or protein complexes on a single-molecule level (8). A more detailed explanation of DNA curtains will follow in **Chapter I 2.**.

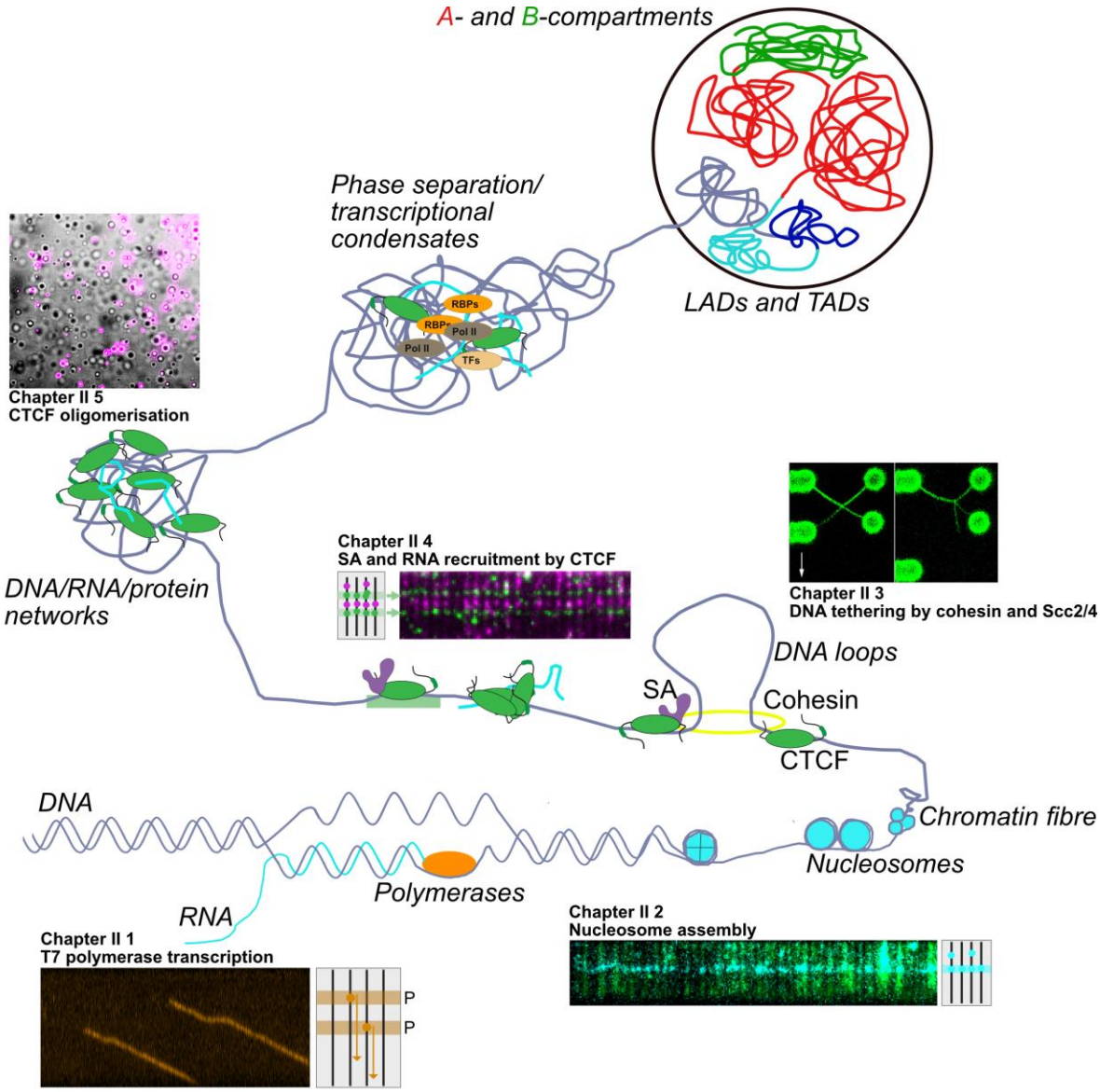



*Figure 6: The variety of genome organisation investigated with DNA curtains. The DNA inside eukaryotic genomes is organised by transcribing polymerases and transcription-produced RNA (Chapter II 1.), nucleosome spacing (Chapter II 2.) and formation of the chromatin fibre. Higher levels of genome organisation include TAD formation by chromatin loops (Chapter II 3.)* (230)*, the*

*formation of DNA/RNA/protein networks (Chapter II 4.), oligomerisation of architectural proteins (Chapter II 5.), phase-separated condensates and finally the division into A- and B-compartments. Figure based on (24).*

As explained above, DNA-bound proteins can act as roadblocks for transcribing polymerases, causing backtracking, pausing or transcription termination (404). Additionally, transcription influences many processes in the nucleus, such as the positions and strength of TAD boundaries, phase separation and the formation of compartments (28, 163, 168, 313). It is hence of great interest to understand what influences the DNA interaction of RNA polymerases on a molecular level. Eukaryotic RNA polymerases are large multi-subunit protein complexes (413), which require cofactors for each transcription step, making single-molecule measurements time-consuming and error-prone. A common model to study transcription *in vitro* is the T7 polymerase, a rapidly transcribing single-subunit bacteriophage RNA polymerase (421). In **Chapter II 1. A high throughput single-molecule assay to study transcription**, I developed a single-molecule transcription assay to analyse DNA binding, IC to EC transition, transcription velocities, transcription termination and collisions between multiple polymerases.

DNA-binding properties of proteins are influenced by the fact that DNA inside the nucleus is not naked but wrapped around nucleosomes (11, 422, 423). This influences DNA accessibility, so chromatinized DNA is hence a more qualified substrate to study proteins involved in transcription, replication and repair (31, 135, 424–429). I produced chromatinized DNA by inserting 19 nucleosome positioning sequences into the DNA used for nucleosome assembly with recombinantly expressed and fluorescently labelled histone octamers. The results are summarised in **Chapter II 2. "Nucleosome assembly on DNA curtains".**

Architectural proteins play a major role in genome organisation. Cohesin shapes TADs by extruding loops in a fast and dynamic fashion (13–15) while also holding sister chromatids together during cell division (20, 235), which requires stable and long-lasting protein-DNA interaction. This raises the question of whether these different kinds of DNA interactions can be visualised on DNA curtains. The study (230) in **Chapter II 3.** analyses cohesin's DNA-binding activity, DNA-tether formation, lifetime and movement on DNA. Additionally, optical tweezers are applied to study cohesin's ability to form bridges in cis (on the same DNA molecule), resembling DNA loops and in trans (between two DNA molecules), resembling sister chromatid cohesion.

The second architectural protein that I was focusing on during my work was the transcriptional insulator CTCF. This protein acts as an anchor of cohesin-mediated looping by recognition of a specific CBS on DNA (18, 19, 204). The study (229) in **Chapter II 4.** uses fluorescently labelled CTCF and truncation mutants to reveal which parts of CTCF are required for CBS recognition. To understand the molecular basis of loop extrusion, combined measurements with SA and CTCF were performed. Additionally, the single-molecule transcription assay established in **Chapter II 1.** was applied with previous CTCF enrichment on its CBS to study how CTCF influences transcription and vice versa. CTCF-RNA cluster formation previously observed *in vitro* (225) was analysed by performing single-step photobleaching experiments of CTCF and analysing interactions with fluorescently labelled RNA.

Condensates are often enriched in RBPs (28), and CTCF was shown to also form condensates (221, 224, 430). In **Chapter II 5. CTCF forms oligomeric structures with and without nucleotides** CTCF's ability to form oligomers was analysed using mass photometry, CTCF condensate formation with an *in vitro* droplet assay and formation of larger clusters with DNA using atomic force microscopy (AFM).

## I 2. Methodological introduction

Single-molecule techniques can be applied to study the molecular mechanisms of biological interactions, by displaying binding dynamics and conformational changes of individual molecules instead of averaging a large number of events, like in bulk measurements (431, 432). This allows the visualisation of individual subunits of large complexes and of conformations which are short-lived or occur less frequently and which therefore would be concealed by averaging (431, 432). Single-molecule measurements often depend on measuring molecular forces (433) for example in the case of optical tweezers (434) and AFM (435), further explained below. They can also depend on optical imaging (433) for example in the case of mass photometry (436) and total internal reflection fluorescence microscopy (TIRFm) (437), as described below, or in the case of super-resolution microscopy including stochastic optical reconstruction microscopy (STORM) and stimulated emission depletion (STED) (438–440).

### 2.1 Mass photometry

A single-molecule technique applied in this study is mass photometry, used for measuring the masses of individual biomolecules in a solution. In mass photometry the binding and unbinding events of these molecules, for example, proteins or nucleic acids, to a glass surface are detected (436) **(Figure 7A)**. Binding events lead to the scattering of laser light, and the contrast between the scattered light and the light reflected from the glass surface is recorded (interferometric contrast) (436, 441). This contrast is directly proportional to the polarizability, which depends on both the volume and refractive index of the studied molecule (441). By using a calibrant of the same biomolecule class (for example, BSA for proteins), assuming the same refractive index, one can quantify the mass of single biomolecules in solution from their volume (436). Mass photometry has been used to determine the domain composition of protein complexes like cohesin or the proteasome (442), the oligomerisation dynamics of actin (443) and nucleic acid binding for Cas9 (444). In **Chapter II 5**., mass photometry was acquired to study the concentration-dependent oligomerisation of CTCF. Mass photometry is, hence, a powerful tool to examine all aspects of protein interactions (436) but does not provide details about molecular mechanisms or conformational changes.

### 2.2 AFM

AFM is an imaging technique that tracks the surface of a biomolecule to provide information about its shape and size (433, 445) **(Figure 7B)**. To obtain structural information, changes in vertical distance are measured by recording the deflection of a laser from a cantilever (433). Different measurement modes exist for AFM imaging. In the contact mode, the AFM-tip is in constant contact with the sample while recording sample height, which can, however, lead to the damaging of soft biological samples (445). This can be circumvented by two different dynamic imaging modes (AC modes) in which the cantilever oscillates (445). In the intermittent mode or tapping mode, changes in oscillation frequency caused by changes in sample height and changes in the phase of the frequency wave depending on sample softness are recorded, significantly reducing sample damage by forming only very short contacts with the sample (445, 446). In the non-contact mode, repulsive and attractive forces from the sample surface lead to changes in oscillation frequency and the tip never actually touches the surface (445). Unlike other common methods in structural biology, like X-ray crystallography and cryogenic electron microscopy, AFM can work in solution under physiological conditions preventing sample preparation artifacts and allowing protein dynamics and reaction kinetics to be visualised (447). Unlike fluorescence microscopy, it does not require labelling of biomolecules, which is time-consuming, requires additional purification steps

and can lead to changes in protein function or folding (447). AFM measurements at ~1 nm resolution (447) can be applied to gain insights into biomolecular structures (448), explore the mechanisms and conformational changes upon DNA, RNA and ssDNA binding of proteins (228, 435, 449) and even study dynamic processes like loop extrusion (450). **In Chapter II 5.**, AFM was applied to visualise how oligomeric CTCF forms large clusters with DNA.

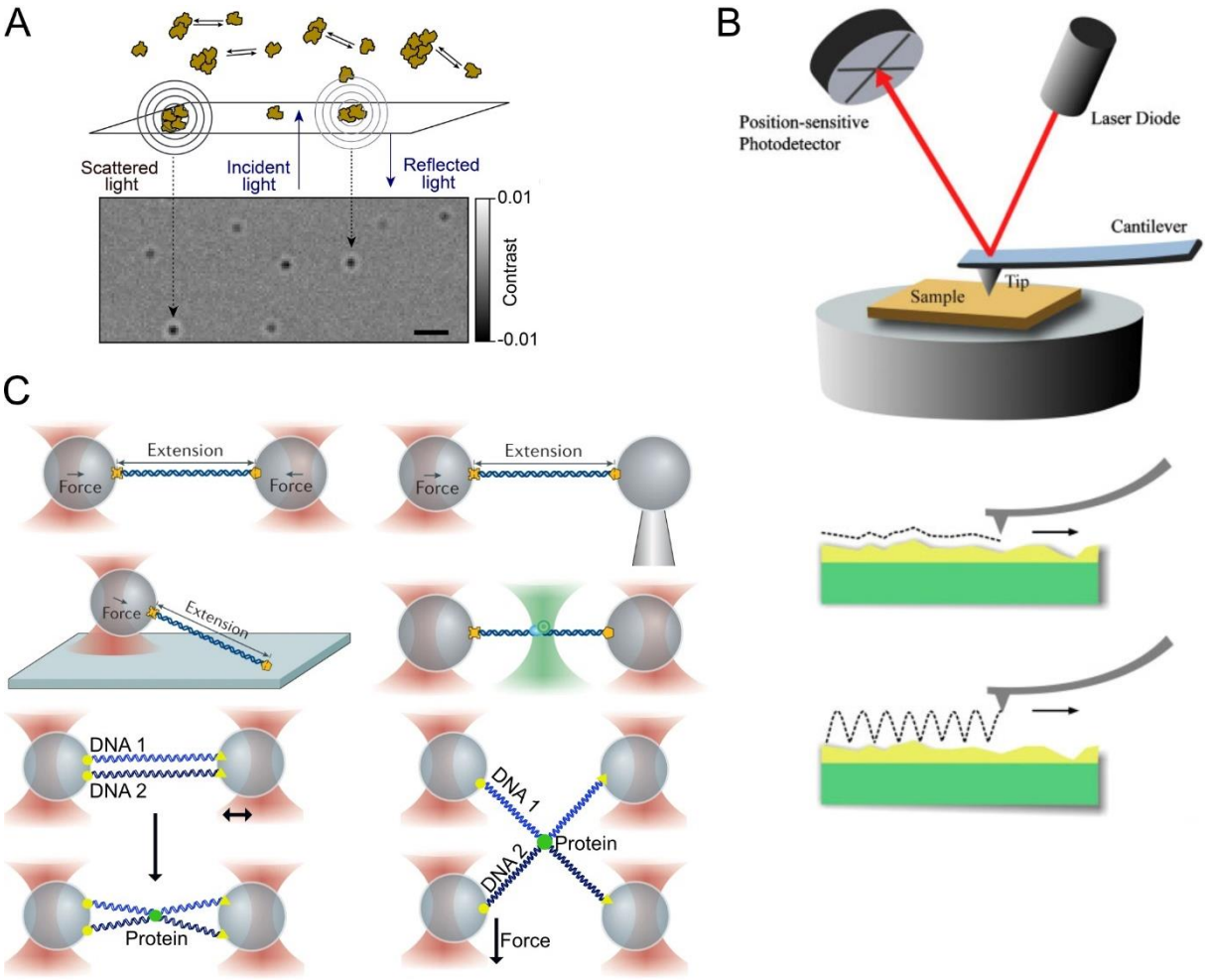



*Figure 7: Different single-molecule methods. A) Mass photometry binding events lead to changes in light scattering caused by single molecules or oligomers binding to and unbinding from the glass surface. Figure from (441). B) Sample structure is obtained during AFM imaging by recording the laser signal deflected from a cantilever raster scanning the sample surface. Different imaging modes are contact mode (top) and dynamic mode (bottom). Figure from (445). C) Optical tweezer experimental setup. A DNA molecule is attached to a micron-sized bead trapped in a laser beam on one end and attached to a second bead (top left), attached to a bead captured by a micropipette (top right), or to the sample surface (2nd from top left) on the other end (451). Combining optical tweezers with confocal microscopy allows simultaneous visualisation and force manipulation of single molecules (2nd from top, right) (451). Two DNAs can be caught simultaneously to test if a protein (green) can bridge DNAs in trans (bottom left). A quadruple trap is used to test the force stability of trans-tethers formed by a protein (green) (bottom right). Figure from (451).*

## 2.3 Optical tweezers

To achieve sub-piconewton resolution, a different force manipulation method called optical tweezers can be applied (451). In dual-trap optical tweezers, micron-sized beads can be trapped within the focus of the laser beams (451). The beads are functionalised to allow the site-specific attachment of biomolecules between the beads in a dumbbell geometry (434, 452) **(Figure 7C top left)**. Other standard setups use a tethering between one trapped bead and either a captured

second bead using a micropipette **(Figure 7C top right)** or the surface (433, 451) **(Figure 7C 2nd from top left)**. These setups can be used to measure force-extension curves to study protein unfolding or to study DNA compaction, as the beads can be pulled together by the interacting molecules (453, 454).

Additionally, by combing optical tweezers with confocal microscopy, binding dynamics and kinetics of fluorescently labelled proteins on the tethered DNA can be tracked **(Figure 7C 2nd from top right)** (451). This was, for example, used to study the force stability of different DNA compaction states caused by a ssDNA-binding protein while also being able to track differences in its diffusion speed at different conformational states (455). It can also be applied to visualise proteins involved in DNA repair by force-dependent opening of DNA strands, creating parts of ssDNA and ssDNA/dsDNA junctions (452). Further optical tweezers can be combined with smFRET, which, for example, revealed that a single UvrD repair helicase can bind to and move on DNA, but that dimerisation is required to generate the molecular force for unwinding large stretches of DNA (456).

Confocal microscopy combined with optical tweezers was applied in **Chapter II 3.** to show that cohesin forms two different kinds of bridges on DNA molecules: force-sensitive reversible bridges and force-resistant permanent bridges. Additionally, two DNAs were trapped **(Figure 7C bottom left)** to analyse the trans-tethering of cohesin. A quadruple trap optical tweezer setup was applied **(Figure 7C bottom right)** to study the stability of formed tethers.

A clear advantage of this method is that studied biomolecules can be simultaneously visualised and force-manipulated (451, 457). Additionally, measurements can be performed at physiological buffer conditions and unlike for TIRFm, measurements are not disturbed by unspecific surface interactions. However, measurements can be time-consuming, and it can be hard to acquire statistically meaningful data.

### 2.4 DNA curtains

In DNA curtain measurements hundreds of DNA-protein interactions can be visualised in one measurement, overcoming the problem of low sample throughput in single-molecule techniques (8). This allows the analysis of protein-binding positions, binding kinetics, movement on DNA, protein oligomerisation and interactions with additional binding partners (8). Additionally, this method is ideally suited to analyse the impact of proteins involved in genome architecture, as protein-DNA interactions can be studied in real-time on a single-molecule level.

DNA curtains use total internal reflection fluorescence microscopy (TIRFm). If laser light travels in a medium with a higher refractive index $n_1$ (e.g. glass) encountering a medium with a lower refractive index $n_2$ (e.g. water or buffer) at an incidence angle above the critical angle $\theta_c$, it is no longer refracted into the second medium but totally reflected (458). $\theta_c$ is given by the following equation:

$$\theta c = \arcsin\left(\frac{n_2}{n_1}\right) \quad (1)$$

When total internal reflection occurs, an electromagnetic wave - called an evanescent wave - is created at the surface, which has the same wavelength as the laser light and exponentially decays in intensity with increasing distance from the surface (458). This means that only fluorophores within ~ 100 nm distance to the surface are excited, while most molecules within the sample chamber are not, which drastically reduces the fluorescent background (458, 459). In the case of DNA curtains, this allows us to visualise single fluorescent molecules interacting with the DNA, which is attached close to the surface (8).

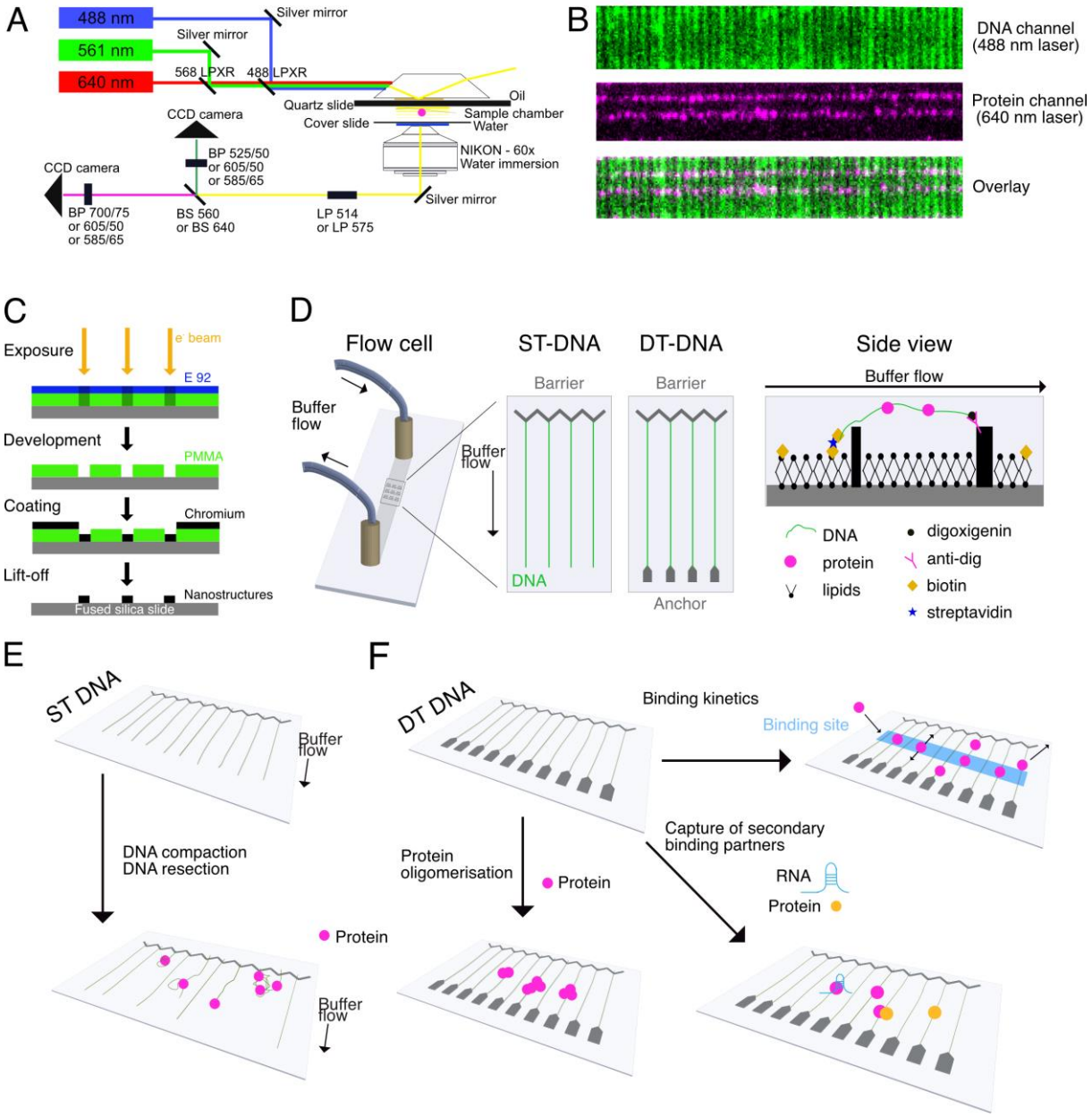



*Figure 8: Overview of the DNA curtains method. A) Scheme of the setup for TIRFm. Three different lasers are aligned using dichroic mirrors and silver mirrors and focused onto a glass prism above the critical angle. Three wavelengths can be used by alternating laser excitation (ALEX), which allows multi-coloured single-molecule experiments. Depending on the used fluorophores and lasers, different long path filters, a 560 nm or 640 nm dichroic mirror and different bandpass filters for red and green channels are used before the light is detected on CCD cameras. B) Example picture of DNA curtains showing YOYO-1-labelled DNA in the 488 nm laser green channel (top), Alexa-Fluor 660-labelled CTCF in the 640 nm laser red channel (centre) and an overlay (bottom). C) Generation of chromium barriers by e-beam lithography. A thin layer of polymethylmethacrylate (PMMA) (green) followed by a thin layer of electra 92 (E92) (blue) is applied on fused silica slides by spin coating. Afterwards, barrier and anchor structures are exposed by the electron beam (grey area within PMMA). The E92 is washed off, and the PMMA is developed. Using an evaporator, a 25 nm chromium layer (black) is applied to create the Chromium-nanostructures before the remaining PMMA is washed off with acetone. Figure adapted from (460). D) Model of DNA curtain measurements. Modified DNA is attached with biotin-streptavidin interactions to a lipid bilayer. By applying a flow, the DNA is pushed against the chromium barriers. The DNA can be double-*

*tethered to the chromium anchor structures separated by 13 µm from barriers, via a digoxigenin-antidigoxigenin interaction. A typical flow cell contains nine measurement areas of multiple barrier and anchor structures each. Measurement solutions can be injected via a microfluidics system. Figure adapted from (460). E) Applications of DNA curtains. Single-tethered (ST) DNA curtains- lacking anchor structures- are applied to measure DNA resection and DNA compaction. F) Double-tethered (DT) DNA curtains are applied to measure protein-binding kinetics, preferred binding sites, protein movements like diffusion or translocation, protein oligomerisation by photobleaching experiments and interaction with secondary protein binding partners like other proteins or RNA.*

In the TIRF setup **(Figure 8A)**, three different light sources (Coherent® OBIS™ 1220123 | 488 nm LX 150 mW Laser; Coherent® OBIS™ 1280720 | 561 nm LS 150 mW Laser; TOPTICA Photonics iBeam smart-S 640-S) are used, which are aligned using two silver mirrors and two dichroic mirrors (488 LPXR and 568 LPXR). The laser light passes through a fused silica prism onto a microfluidic flow cell, at an angle at which it is totally reflected to generate an evanescent wave (8). Emitted photons pass through a microscope objective (Nikon Ti2-e), with a 60x magnification, which can be combined with a 1.5x magnification lens. Depending on the measurement mode, an LP 514 filter to block 488 nm laser light or an LP 575 filter to block both 488 nm and 561 nm laser light is used. Light is split into two different channels by a 560 nm or 640 nm dichroic mirror and further filtered using 525/50 nm, 585/65 nm, 605/50 nm or 700/75 nm filters, depending on the used fluorophores. Finally, CCD cameras (iXon Life 897 ANDOR) are used for detection. An example picture of DNA curtains measurements with fluorescently labelled DNA (top), CTCF binding to two 4 x CBSs (centre) and an overlay of both channels (bottom) is given in **(Figure 8B).**

The DNA curtains approach relies on the 48.5 kbp bacteriophage λ-genome (461). This has several advantages: λ-DNA possesses sticky ends (12 bp overhangs), which can be used to attach biotin and digoxigenin modifications via short complementary oligomers (461, 462). Additionally, λ-DNA allows the insertion of specific DNA sequences, like protein-binding sites, into different sites of the λ-genome by restriction enzyme digestion and subsequent ligation (463, 464). In this work, I used two cloning sites to insert T7 promoter sequences, CBSs or nucleosome positioning sequences, positioned between NgoMIV (20040) and XbaI (24508) or XhoI (33498) and NheI (34679) restriction enzyme sites. The modified λ-DNA needs to be between 38 and 52 kbp in size to produce viable molecular hybrids that can be packed into λ-phages (463–465), which then can be used to infect *Escherichia coli (E. coli)* strains (463, 464, 466) for DNA amplification and subsequent purification.

Besides illumination and DNA constructs, the DNA curtains approach relies on a microfluidic flow cell, which contains chromium-nanostructures as diffusion barriers and anchor structures for the alignment of DNA. To manufacture these flow cells, multiple preparation steps are required. First, two holes are drilled into fused silica slides, which are later required for the application of measurement solutions via a microfluidic system. After the removal of organic material using peroxymonosulfuric acid, nanostructures are created using e-beam lithography (EBL). A thin layer of PMMA, topped by a thin layer of the conductive resist E92, is spin-coated on the slides **(Figure 8C)**. Next, an electron beam is used to write nanostructures into the coating. The exposed pattern is further developed using methyl isobutyl ketone (MIBK) to create gaps in the coating of the slides, which creates a mask for the deposition of a 25 nm chromium layer. Finally, PMMA is removed by an acetone wash, and only the chromium-nanostructures remain on the fused silica slide. A measurement channel is formed by placing double-sided tape around the chromium-nanostructures, which are overlaid by a cover slip **(Figure 8D)**. Nanoports are glued to the holes on the opposite side and connected to outlet tubing. Inside the measurement channel are 3x3 measurement areas containing the chromium-nanostructures.

A main disadvantage is the PMMA autofluorescence and the scattering of the excitation beam by the chromium-nanostructures, as observed in our case for the 640 nm laser, reducing the signal-

to-noise ratio (467). An alternative to chromium barriers is hydrogen silsesquioxane (HSQ). It can be lithographically patterned directly on fused silica slides, which makes them easier to produce than chromium barriers and is transparent across the ultraviolet to infrared spectrum, which is a clear advantage over scattering metallic barriers (468). Both chromium and HSQ structures can be reused by applying multiple wash steps to remove surface-bound proteins, lipids, DNA, RNA and glue from previous experiments. However, chromium-nanostructures can become damaged over time causing holes in the diffusion barriers and reducing the tethering efficiency to chromium anchors. Furthermore, organic remnants often stick to the glass surface and can negatively impact further measurements. Unlike chromium-nanostructures, for which we use ethanol, acids and bases, HSQ structures are thereby damaged in even mild solvents (467). As a more resistant alternative nano-trench patterns can be used, which apply a combination of EBL and etching to create sawtooth-patterned holes in the glass surface as diffusion barriers (467). They have the additional advantage of avoiding PMMA autofluorescence and the scattering of chromium-nanostructures (467). However, DNA molecules need to be stretched by constant buffer flow when only using a single diffusion barrier. This leads to high sample consumption and the requirement of a constant force applied to the studied molecules during visualisation (469), which might influence their binding dynamics and movement on DNA. A clear advantage of using a combination of chromium diffusion barrier and anchor structures is the ability to permanently stretch the DNA in the absence of buffer flow. Additionally, this allows us to study many protein properties that cannot be visualised when just using a single diffusion barrier, which I will explain in further detail below.

For DNA curtain measurements the λ-DNA is immobilised on the flow cell surface via attachment to a lipid bilayer. The lipids are in part biotinylated, which interacts with streptavidin and enables the binding of biotinylated DNA **(Figure 8D)**. The lipids function in surface passivation, reducing hydrophobic interactions between the studied molecules and the glass surface (470, 471). In contrast to other methods like polyethylene glycol (PEG) passivation (472) or using proteins like bovine serum albumin (BSA) or casein (473, 474), lipid bilayers are fluid and allow diffusion of attached molecules (8, 470, 471). Slides for single-tethered (ST) DNA curtains possess “zic-zac”-shaped chromium-nanostructures **(Figure 8D,E)**, which form separated wells for the DNA molecules (8, 460, 475) and serve as diffusion barriers. In case a hydrodynamic force, like a buffer flow, is applied all lipid-attached DNAs will move in the same direction (8). DNA can, therefore, be pushed against the barriers, which stop the movement of the DNA (8). This allows simultaneous visualisation of multiple DNA strands, which are aligned parallelly and are all in the same orientation as the biotin-attached DNA end is positioned inside the well (8, 460, 471). To visualise protein-DNA interactions in the absence of buffer flow, DT DNA curtains can be used, which contain additional chromium-nanostructures called anchors, spaced 13 µm from the barriers **(Figure 8D)**. Here, DNA is attached on one side to the lipids and pushed against the diffusion barriers, while the opposing end is additionally attached to the anchor structures **(Figure 8D,F)**. Double-tethering requires an incubation with anti-digoxigenin (anti-dig), which binds unspecifically to chromium anchor structures **(Figure 8D)** and specifically recognises DNA, which is end-modified with digoxigenin. Small gaps between anchor structures ensure that no DNA is trapped between the barrier and anchor structures. The 48.5 kbp bacteriophage λ-genome has a contour length of 16.5 µm (assuming 0.34 nm per base pair) (476), meaning that it is stretched to about 79% of its length in DT DNA curtains measurements.

On DNA curtains, the exact binding positions of proteins can be analysed. This can be used to identify position-specific binding depending, for example, on AT-ratios (197, 229), on specific sequence motifs (229, 477–479) or the presence of DNA secondary structures, like R-loops (201, 480) and sites of DNA damage (452, 469, 479, 481, 482). For ST curtains measurements, DNA is only attached on one end, which allows the visualisation of certain properties of DNA-binding proteins like the ability to compact (22, 329, 483, 484) or to resect DNA (485) **(Figure 8E)**. On double-tethered (DT) DNA curtains, bound molecules can be tracked to study the translocation of

motor proteins (229, 479, 486, 487) or to study the one-dimensional (1D)-diffusion behaviour of proteins along the DNA strand at different salt concentrations (197, 230, 480–482). Additionally, binding and unbinding events can be observed in real-time to study binding kinetics on different sequences and at varying buffer conditions (197, 201, 230, 478, 487) **(Figure 8F)**. Collisions and interactions with other DNA-bound proteins can be visualised (197, 469, 479, 482). Protein oligomerisation on DNA can be verified by photobleaching (229, 478) or dual-colour experiments (452). Additional DNA, RNA or other proteins can be applied via the microfluidic system to analyse possible interactions with already DNA-bound proteins (201, 229, 452). This makes DNA curtains an ideal tool to study the proteins introduced in **Chapter I 1.** and the properties and DNA-binding mechanisms underlying their various roles in genome architecture and transcription, like CTCF sequence recognition and lifetime on DNA, SA diffusion, T7 polymerase translocation, or tethering of two DNAs by cohesin.

# II Results

## II 1. A high throughput single-molecule assay to study transcription

**Summary:**
T7 polymerase is a rapidly transcribing single-subunit RNA polymerase, frequently used for recombinant gene expression. Due to its importance for biological research as well as the biotechnological industry, its properties have been extensively studied for many years, revealing different conformational states during transcription initiation, elongation and termination and the presence of different promoter, pause and termination sequences. However many details remain unknown, for example, how different DNA sequences as well as different buffer conditions influence successful transcription initiation as well as velocity and processivity during transcriptional elongation.

In this study, I use a high-throughput single-molecule assay to analyse initiation to elongation complex transition, transcription velocity, processivity and collisions between multiple polymerases. Here, I display that lowering nucleotide concentrations leads to lower transcription velocities and increases the fraction of not starting polymerases. There have been mixed results on whether DMSO (dimethyl sulfoxide), a common buffer additive for *in vitro* transcription, enhances transcription. I show that it slightly reduces transcription velocities but leads to a reduction of not starting polymerases, showing that it can increase the efficiency of *in vitro* transcription reactions.

Additionally, I use promoters with two different transcription initiation but identical promoter binding sequences. I thereby show that promoter sequence has an impact on the fraction of promoter unbinding vs. successful transcription initiation by influencing the stability of promoter-bound T7. I show that DNA sequence further influences T7 polymerase during elongation since T-rich sequences lead to an increase in transcription termination. This can be used to optimise DNA sequences used for recombinant gene expression for efficient T7 transcription.

I then utilised this behaviour to study the collisions of transcribing polymerases on DNA, showing that actively transcribing T7 polymerase can push stalled polymerases off or along the DNA, revealing a mechanism of how polymerases can clear the DNA of stalled complexes at highly transcribed genes.

This study therefore reveals how DNA sequence and buffer conditions influence the different stages of T7 transcription and how polymerases can deal with stalled complexes.

**Author contributions:**
Joelle-Deplazes Lauber and Sarah Zernia performed protein expression, purification and labelling. Jonas Huber carried out DNA curtains measurements. Johannes Stigler and Jonas Huber analysed DNA curtains measurements. Jonas Huber wrote the manuscript with input from all other authors.

**Abstract:**
T7 polymerase is an enzyme frequently used in the biotechnological and pharmaceutical industry as well as biological research since it only consists of a single subunit and allows for fast and highly processive transcription. To optimise recombinant gene expression, it is therefore of exceeding importance to analyse the mechanism of T7-transcription steps as well as the influence of reaction and buffer conditions. Since T7 is used for the expression of a large number of different genes the influence of DNA sequence on transcription was analysed using a single-molecule assay called DNA curtains, allowing simultaneous visualisation of multiple DNA-protein interactions. High T-content led to increased polymerase unbinding during transcription initiation and impaired T7 processivity. Lowering nucleotide concentrations also had a negative influence on successful transcription initiation, while DMSO, a common buffer additive, had a positive effect. A dual-colour labelling approach revealed how polymerases can clear genes by pushing stalled complexes off or along the DNA.

## *1.1 Introduction*

Highly efficient recombinant gene expression is required for both biochemical characterisation in life science as well as for medical and agricultural applications in industrial processes (488). *E. coli* is often used as a host for recombinant gene expression due to the simplicity and low costs of growth conditions and its well-characterised genetics (489). T7 polymerase-based vectors are thereby most frequently used (489) since they can be applied as an orthogonal expression system in *E. coli*, meaning that *E. coli* polymerases do not transcribe any genes in the control of the T7 promoter and T7 polymerase does also not transcribe *E. coli* genes (490).

T7 polymerase is a single-subunit RNA polymerase without the requirement for any cofactors for efficient transcription (421, 491), transcribing the largest part of the viral genome during *E. coli* infection (491–494). T7 polymerase consists of an N-terminal domain and a polymerase domain, the second being further divided into thumb, finger and palm (495, 496) **(Figure 1A)**. T7 promoters consist of two motifs, an upstream binding region (positions −17 to −5), bound by the T7 polymerase N-terminal domain and specificity loop and a downstream initiation region (-4 to +6) initially opened up by the enzyme at positions -4 to -1, forming the initiation complex (IC) (496–500) **(Figure 1B)**. The template strand is positioned in a deep and highly positively charged pocket and RNA synthesis starts by binding of a GTP at position +1 (496, 498, 501).

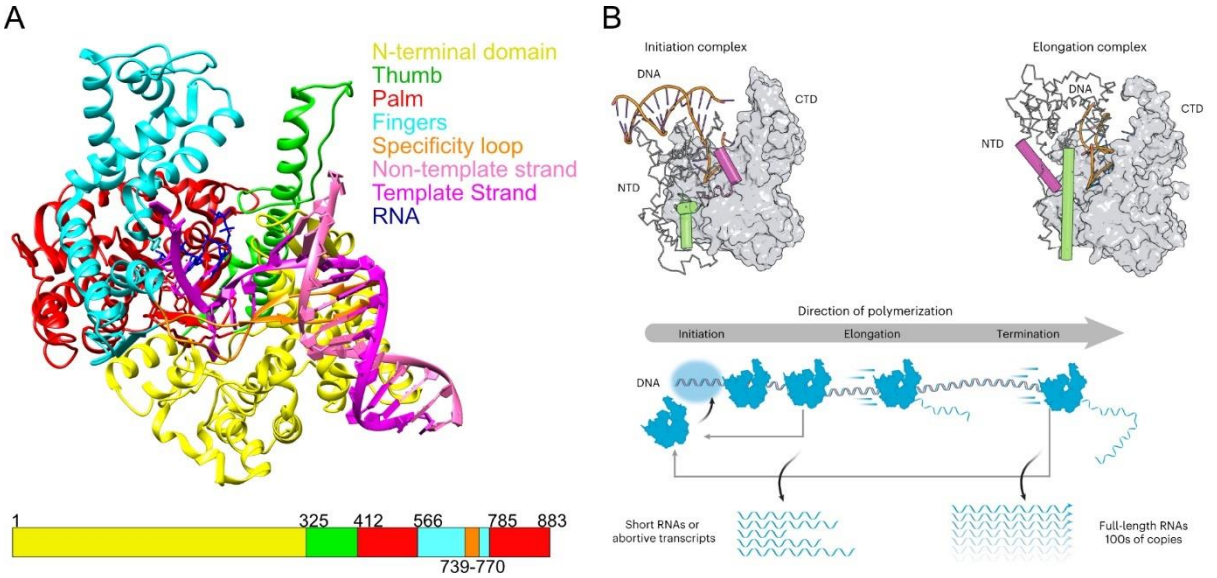


*Figure 1: Structure and transcription cycle of the T7 polymerase. A) X-ray structure of a T7 polymerase IC (PDB: 1QLN) with individual subunits shown in the respective colours. The N-terminal domain and specificity loop are required for the interaction with the upstream promoter binding sequence and the palm, fingers, thumb, N-terminal domain and specificity loop all form the positively charged active site containing the melted DNA and three bases of RNA. Structure from (496). B) During transcription T7 polymerase binds the upstream part of the promoter DNA in multiple cycles and produces short abortive transcripts in the less stable IC conformation, while a conformational change leading to the transition to the more stable EC conformation allows highly processive transcription elongation until a termination sequence is reached. Figure from (499).*

The IC forms less stable transcripts than the elongation complex (EC), which leads to the production of short RNAs released from the T7 polymerase during the early (2-8 nt) and late initiation phase (9-12/13 nt) with subsequent transcription restart, called abortive cycling (379, 380) **(Figure 1B)**. A specificity loop of T7 polymerase **(Figure 1A),** positioned between finger and palm subdomains is involved in promoter recognition in the IC and after a conformational change in stabilising the EC presumably by binding to RNA transcripts at an RNA length of 9 nucleotides (498, 500). The N-terminal domain also undergoes large conformation changes during the transition to the EC, no longer binding to upstream promoter DNA and allowing the formation of an RNA exit channel (491, 502–504).

The more processive EC is then able to produce long RNAs until reaching a termination sequence (499, 501). There are two types of termination signals for T7 polymerase. Class I can be found in the viral genome and depends on stem-loop formation followed by further destabilisation via multiple Us, while for class II, a direction-dependent 7 bp conserved sequence causes pausing, which in the presence of additionally destabilising T-rich sequences causes unbinding (505–512). Since Class II causes frequent polymerase pausing (505, 510–513) complexes paused at Class II sites could resemble roadblocks for trailing elongating polymerases. In *E. coli* it has been shown that multiple polymerases can cooperate to overcome pausing and roadblocks (402, 409, 514). However, whether such a mechanism also exists for T7 polymerase or if instead colliding polymerases inhibit transcription remains unclear.

Due to its simplicity, a large amount of technologies (501) including for example DNA-sequencing (515), RNA-sequencing (516), vaccine production (517) and protein production by coupling to *in vitro* translation (518) use the T7 polymerase system and therefore depend on optimised T7 polymerase reaction conditions. The impact of reaction components like nucleotide concentration, pH, NaCl and $MgCl_2$ concentration and several buffer additives like spermidine, DMSO or different non-ionic detergents have been extensively studied both in bulk (519) and to some extent also in single-molecule experiments (520, 521). Decreasing nucleotide concentrations leads to a decrease in transcription speeds (520, 521) and leads to an increased time spent in the IC (382). Increased RNA production by T7 polymerase was observed in the presence of up to 20 % DMSO (522), however, other studies found no influence of DMSO on transcription (519, 523). How the different stages of transcription initiation, elongation and termination are influenced by DMSO remains unclear.

Most T7 promoters contain a common binding sequence from position -17 to -5, however, the promoters can differ in the initiation sequence from -4 to +6 nucleotide position (497, 498). It has been shown that sequence differences in the initiation region can influence transcription efficiency (497, 501, 519), which could be a cause of multiple mechanisms including slowed IC to EC transition, more frequent RNA release and unbinding from promoter DNA (379, 380, 382, 524, 525).

Here, I apply DNA curtains to study the influence of two different promoter sequences on transcription initiation. I show that while T7 polymerases bind to both promoters, efficient transcription initiation is only observed for one initiation sequence, while the other causes frequent

polymerase unbinding. I display that this is caused by a decreased polymerase lifetime on the second initiation sequence. Additionally, I show that DNA sequence also impacts transcription elongation with T-rich sequences leading to more frequent pausing and unbinding. I use a dual-colour single-molecule assay to study collisions between paused polymerases and differently labelled trailing polymerases. Trailing polymerases most often stop when colliding with a paused leading polymerase but can also push leading polymerases off or along the DNA revealing a potential mechanism for efficient transcription by clearing genes of stalled proteins. I also analysed buffer conditions to display that while DMSO has only a marginal influence on transcribing polymerases, it increases the efficiency of the IC to EC transition, causing less unbinding of T7 polymerases.

### *1.2 Material and methods*

**Protein expression and purification**

T7 polymerase was cloned from the *Escherichia phage T7 (Bacteriophage T7)* genome (bp 3171-5822) with an N-terminal 6xHis and a ybbR-tag into a pQE-30 expression vector. T7 polymerase was expressed in *E. coli* LOBSTR for 3 h at 37 °C and the expression was induced between OD 0.4 and 0.6 with 0.5 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). Cells were harvested by 10 min centrifugation at 3500 rotations per minute (rpm) and 4 °C, followed by resuspension in residual LB and a second centrifugation at 3900 rpm and 4 °C for 10 min. Cells from 1 L expression were resuspended in 35 ml lysis buffer (50 mM NaPi pH 8.0, 300 mM NaCl, 1 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF)) and 6 mg/ml lysozyme (Carl Roth #8259.1) was added. Cells were sonicated for 3 x 6 min at 50 % on-rate and 45 % amplitude on ice before centrifugation at 15000 g for 30 min at 4 °C. The supernatant was applied to a gravity column containing 4 ml Ni-NTA slurry (Macherey-Nagel #745400.100). The column was washed with 10 column volumes (CV) nickel wash buffer (50 mM NaPO pH 8.0, 300 mM NaCl, 10 % glycerol, 10 mM imidazole, 1 mM DTT, 20 µg/ml PMSF). The protein was eluted in 2 x 2.5 ml fractions using nickel elution buffer (50 mM NaPi pH 8.0, 300 mM NaCl, 10 % glycerol, 200 mM imidazole, 1 mM DTT). Afterwards, protein-containing fractions were pooled and concentrated to 600 µl by centrifugation at 4 °C and 3900 rpm with Amicon filters with a 30 kilodalton (kDa) molecular weight cut-off (MWCO).

For ybbr-tag labelling, Sfp (made in house, plasmid was kindly provided by the Gaub-lab, LMU Munich) was mixed with T7 polymerase in a molar ratio of 1:1 and a 1.25 excess of LD555- or LD655-CoA dye (Lumidyne) as well as 10 mM $MgCl_2$ in a total volume of 600 µl. Labelling was carried out at 4 °C o/n followed by centrifugation at 10 krpm and 4 °C for 10 minutes to remove precipitated proteins. Proteins were then further purified on a Superose 6 column in T7 polymerase buffer (50 mM NaPO pH 8.0, 300 mM NaCl, 10 % glycerol, 1 mM DTT). Peak fractions containing fluorescently labelled protein were pooled, frozen in liquid $N_2$ and stored at -80 °C.

**λ-DNA constructs for DNA curtains**

Wild-type λ-DNA was purchased from NEB. For generation of 2x T7-4xCBSs λ-DNA, a cassette containing one of two different T7 promoter sites (5'-TAATACGACTCACTATAGGAGGA-3' and 5'-TAATACGACTCACTATAGGTGTG-3') were cloned at two positions into λ-DNA using NgoMIV/XbaI and XhoI/NheI sites. The ligation product was then packaged, amplified, functionalised with biotin and digoxigenin and finally purified via size-exclusion chromatography according to a previous protocol (229).

**Single-molecule DNA curtains experiments**

Flow cell preparation for DNA curtains measurements was performed as described previously (229). Single-molecule measurements were performed in T7 measurement buffer (40 mM Tris

pH 7.5, 1 mg/ml BSA, 6 mM $MgCl_2$, 10 mM DTT, 2 mM spermidin) including an oxygen scavenger system containing glucose-oxidase (Carl Roth), catalase (Sigma Aldrich) and 0.8 % glucose. To this end 3.8 nM LD555 fluorescently labelled polymerase was incubated for 5 minutes at 40 °C in the presence of 200 µM or 1 mM ATP, CTP, GTP and UTP in T7 measurement buffer and then directly loaded onto the flow cell. If indicated 10 % DMSO was added to the reactions. NIS Elements (Nikon) was used for video recording and Igor Pro 8 (Wavemetrics) custom-written code for analysis. To visualise the polymerase, a 561 nm laser (140 mW, 0.98 µW/µm$^2$) with 50 ms or 100 ms illumination times and frame delays of 52 ms or 102 ms, respectively, was used. Videos were recorded for 15 minutes at 1 mM and 30 minutes at 200 µM nucleotide concentration. In the case of dual-colour experiments, 3.8 nM of both LD555- and LD655-labeled T7 polymerase was added to the reaction mix and videos were recorded for 30 minutes using alternative single frame light excitation of 640 nm and 561 nm lasers at 140 mW (0.98 µW/µm$^2$) with 100 ms illumination times and 376 ms frame delays.

**Data analysis**

To detect T7 polymerase binding and transcription events fluorophores were localised and tracked as described (526, 527). This was used to analyse run lengths (distance from T7 polymerase binding position to position of unbinding or final stop). Run lengths were plotted in a histogram containing 40 bins.

To correct velocities for pauses, transcription events were divided into pause and transcription segments using a Markov chain Monte Carlo (MCMC) algorithm and mean velocities of individual transcription segments were then plotted in a histogram containing 100 bins and fitted with a double Gaussian.

The fraction of not starting T7 polymerases (T7 polymerases just consisting of a single pause at the T7 promoter site) was determined. The overall amount of promoter-binding polymerases was counted and normalised to 1 nM concentration.

Lifetimes were analysed using a single or double exponential decay according to formulas (1) and (2) with $\tau_1 < \tau_2$. The fraction of short-lifetime polymerases $F_s$ was calculated according to formula (3).

$$y = Ae^{\frac{-x}{\tau}} \quad (1)$$

$$y = A_1 e^{\frac{-x}{\tau_1}} + A_2 e^{\frac{-x}{\tau_2}} \quad (2)$$

$$F_s = \frac{A_1}{A_1 + A_2} \quad (3)$$

For dual-colour experiments kymogram data was analysed to detect collisions between paused leading polymerases and a transcribing trailing polymerase with a different fluorescent label and collisions were counted for cases in which the difference in X- and Y-position during an observed colocalisation event were smaller than one point-spread-function (PSF) 215 nm or 1.21 pixel (pxl). Three different events for paused polymerases (pushed off DNA, pushed along the DNA, or remaining at pause site for at least 50 sec or until tether rupture/combined unbinding/secondary collision event/end of video) and two different events for transcribing polymerases (continuing transcription or stopping at the position of the leading polymerase for at least 50 sec or until tether rupture/combined unbinding/secondary collision event/end of video) were counted.

*1.3 Results*

**The transcription initiation sequence determines transcription efficiency**

While T7 starts from a common binding sequence (498) at position -17 to -5, there are differences in transcription efficiency depending on initiation sequences (497, 501, 519). I therefore set out to investigate the influence of different initiation sequences on T7 transcription on a single-molecule level using a high-throughput TIRF-based method called DNA curtains. I analysed individual fluorescently labelled T7 polymerases transcribing from two promoter sequences (P1 and P2), which I cloned into two different positions of the λ-DNA used for DNA curtains measurements. The two promoters contain the identical T7 polymerase binding sequence. However, there is a difference in transcription initiation sequence with 5'-TATA**G**GTGTG-3' for P1 and 5'-TATA**G**GAGGA-3' for P2 respectively (+1 codon marked in bold).

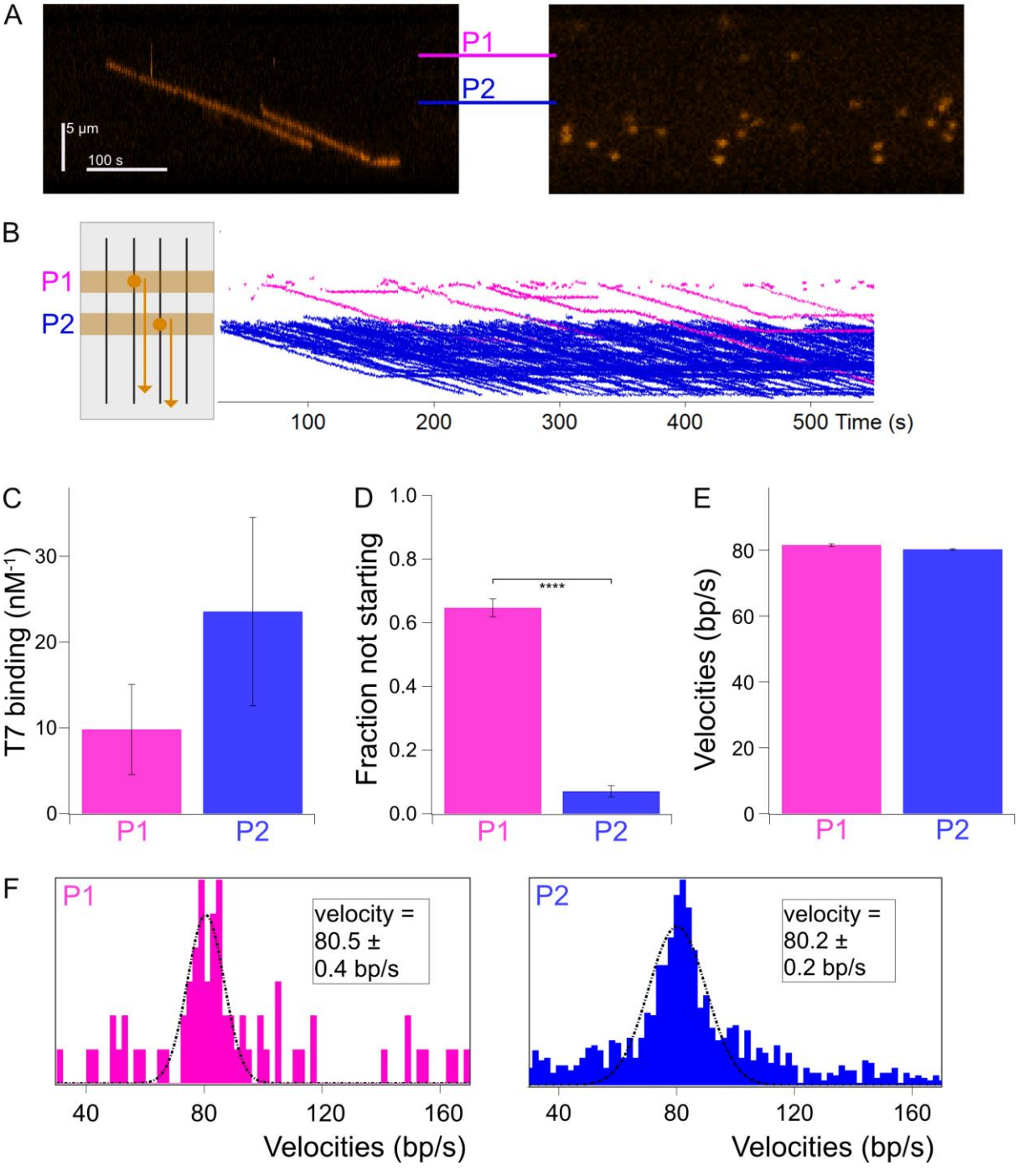


*Figure 2: Promoter initiation sequence determines transcription efficiency. A) Left: Kymograms of representative T7 polymerase transcription events. One T7 polymerase (orange) is starting from promoter P1 and one from promoter P2. Right: Representative microscope image of T7 polymerases. More polymerases are transcribing from the lower promoter P2. B) All tracks of one transcription experiment. Transcription events starting from P1 are marked in pink, and transcription events starting from P2 are marked in blue. C) Amount of binding events normalised to 1 nM concentration for P1 (pink) and P2 (blue) (p = 0.30, N = 5, t-test). D) Fraction of binding but not starting polymerases for P1 (pink) and P2 (blue) (p < 0.00001, N=5, t-test). E) Bar graph showing transcription velocities for P1 (pink) and P2 (blue) (p = 0.53, N = 217 for P1 N = 2281 for P2, z-test). F) Histograms of transcription velocities for P1 (left, pink) and P2 (right blue) displaying the Gaussian fit for transcribing polymerases.*

I calculated kymograms of individual DNA strands **(Figure 2A, left)** and observed individual T7 polymerases binding to the upper promoter P1 or the lower promoter P2 and moving along the DNA in a directed manner. I interpret this movement as active transcription. I recognised that many more transcription events started from the lower promoter P2 compared to P1 **(Figure 2A, right and 2B)**. I quantified the number of binding events and the fraction of polymerases binding without starting transcription and found that 2.4 times more polymerases bind to P2 compared to P1 with however no statistically significant difference **(Figure 2C)** ($p = 0.30$). Strikingly, on P1 more than 60 % of the polymerases failed to start transcribing on DNA **(Figure 2D)**, which is 9 times more than at P2 ($p < 0.00001$). However, once the polymerase started, transcription velocities were comparable for both promoters **(Figure 2E+F)** ($p = 0.53$).

Our results suggest that this difference in transcription initiation sequence causes an alteration in transcription initiation efficiency, which dramatically reduces the number of processive transcription events. This shows how minor changes in the DNA sequence influence the performance of the T7 polymerase.

**Low transcription efficiency is caused by low stability on the initiation sequence**

To get a better understanding of what causes the difference in the amount of transcription events I analysed lifetimes of bound polymerases. The survival plots **(Figure 3A)** displayed a double exponential decay for both promoters. This means that there are two kinds of polymerase species, one falling off very quickly after binding (short lifetime) and another one staying bound (long lifetime). The number of polymerases with a short lifetime was much higher for P1 than for P2 **(Figure 3B**, $p < 0.00001$).

Furthermore, the average short lifetime was significantly reduced at P1 compared to P2 **(Figure 3C**, $p = 0.0002$). In contrast, no significant difference in long lifetimes was observed **(Figure 3D,** $p = 0.82$), showing that the initiation sequence has no influence on the lifetimes of transcribing polymerases.

Our results show that the reduction in starting polymerases is caused by a reduced lifetime of the polymerases on the unfavoured DNA sequence. I assume that the reduced stability increases the chance of polymerase unbinding before having time to complete the IC to EC transition (379, 382). A possible explanation could be that P1 contains more Ts which might lead to more unstable base-pairing between uracil-RNA and DNA (528).

I conclude that the lower stability of polymerases on P1 causes a decrease in the amount of actively transcribing polymerases but has no influence on polymerase lifetime or velocity once transcription has successfully started.

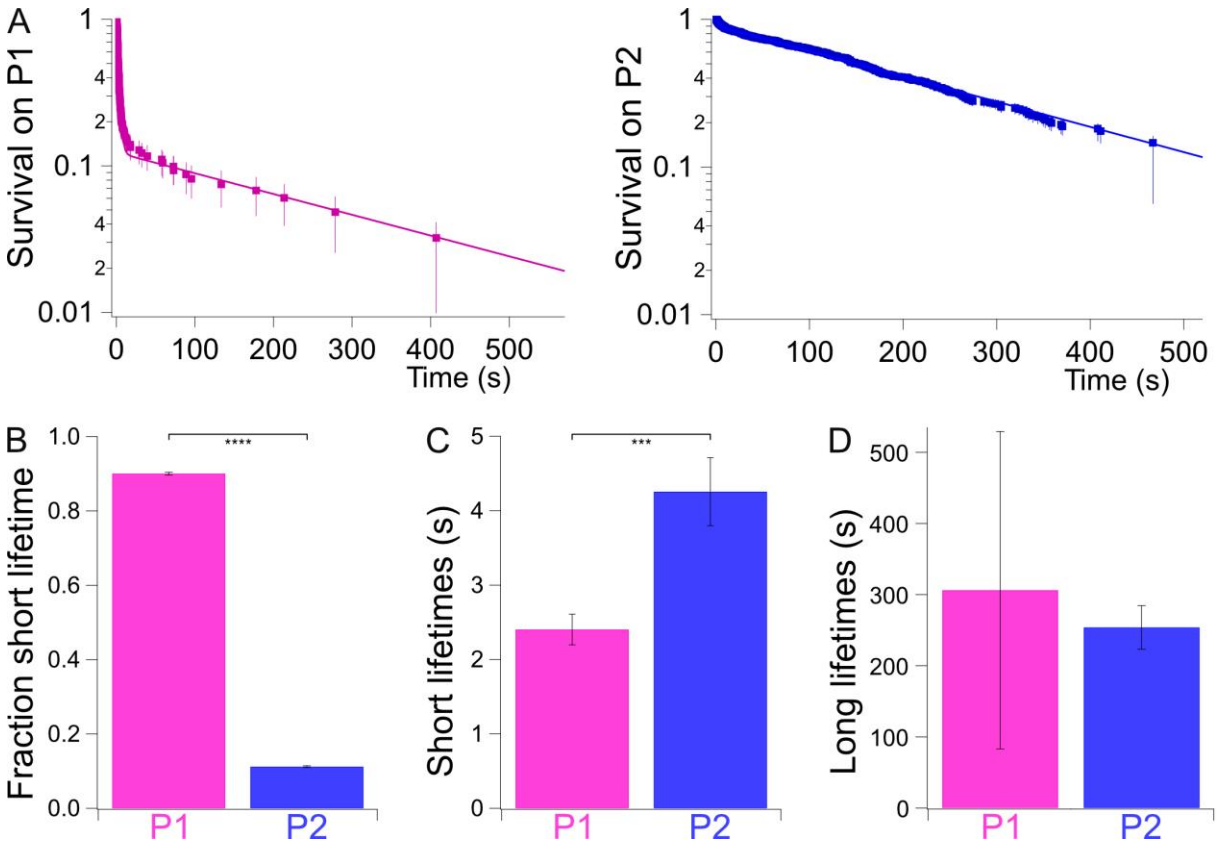



*Figure 3: Increased lifetime at promoter P2 causes higher transcription efficiency. A) Survival plots of polymerases binding to P1 (left, pink N =193) and P2 (right, blue, N=583) fitted with a double exponential decay. B) Fraction of polymerases displaying a short lifetime in the double exponential fit from A) for P1 (left, pink) and P2 (right, blue) ($p < 0.00001$, z-test). C) Short lifetimes of the double exponential fit from A) for P1 (left, pink) and P2 (right, blue) ($p = 0.0002$, z-test). D) Long lifetimes of the double exponential fit from A) for P1 (left, pink) and P2 (right, blue) ($p = 0.82$, z-test).*

**DMSO and nucleotide concentration influence the initiation-elongation transition of T7 polymerase**

As T7 polymerase is commonly used in both industry and research (491), optimisation of T7 buffer conditions was analysed in bulk (519) and, to some extent, also in single-molecule (520, 521). However, whether DMSO does (522) or does not (519, 523) change transcription efficiency is still not clear.

I next wondered how common alterations in polymerase buffer conditions influence polymerase performance in the single-molecule transcription assay. I therefore added either 10 % DMSO in the reaction buffer or reduced the nucleotide concentration by 5-fold (200 µM instead of 1 mM). As the amount of transcription events on P1 is rare, I thereby focused the analysis on transcription events starting from P2.

The addition of DMSO led to a slight but significant decrease in transcription velocities **(Figure 4A)** ($p < 0.00001$), while lowering the nucleotide concentration, as expected, drastically decreased transcription velocities ($p < 0.00001$). This also shows that the polymerases that I investigate here are indeed actively transcribing complexes that require nucleotides for their directed motion.

I next analysed how the transcription initiation is influenced by buffer additives. DMSO had a positive impact, as the amount of non-starting polymerases was significantly reduced to less than

2 % of all binding events **(Figure 4B**, p = 0.04**)**. Since, in the case of DMSO, almost all polymerases start transcribing, I consequently also observed a survival plot with a single exponential decay, displaying no short lifetime fraction **(Figure 4C)**.

In contrast, lower nucleotide concentrations led to a 3.5-fold increase (p = 0.03) in polymerase unbinding **(Figure 4B)**, possibly prolonging IC to EC transition (379, 382) due to slower velocities and, therefore, increasing the chance of unbinding during initiation. Consequently, I also found a higher (p < 0.00001) fraction of short lifetime polymerases **(Figure 4C,D).** These displayed a higher (p < 0.00001) lifetime than at standard nucleotide concentrations **(Figure 4E)**, which agrees with the theory of prolonged IC to EC transition by slower nucleotide addition rates. In all three cases, the lifetimes of transcribing polymerases are around 4-5 min **(Figure 4F)**.

I conclude that lower nucleotide concentrations decrease transcription efficiency by causing a decrease in velocity and successful transcription initiation. In contrast, DMSO leads to a slight reduction in velocities but increases transcription efficiency by almost completely preventing unbinding during initiation.

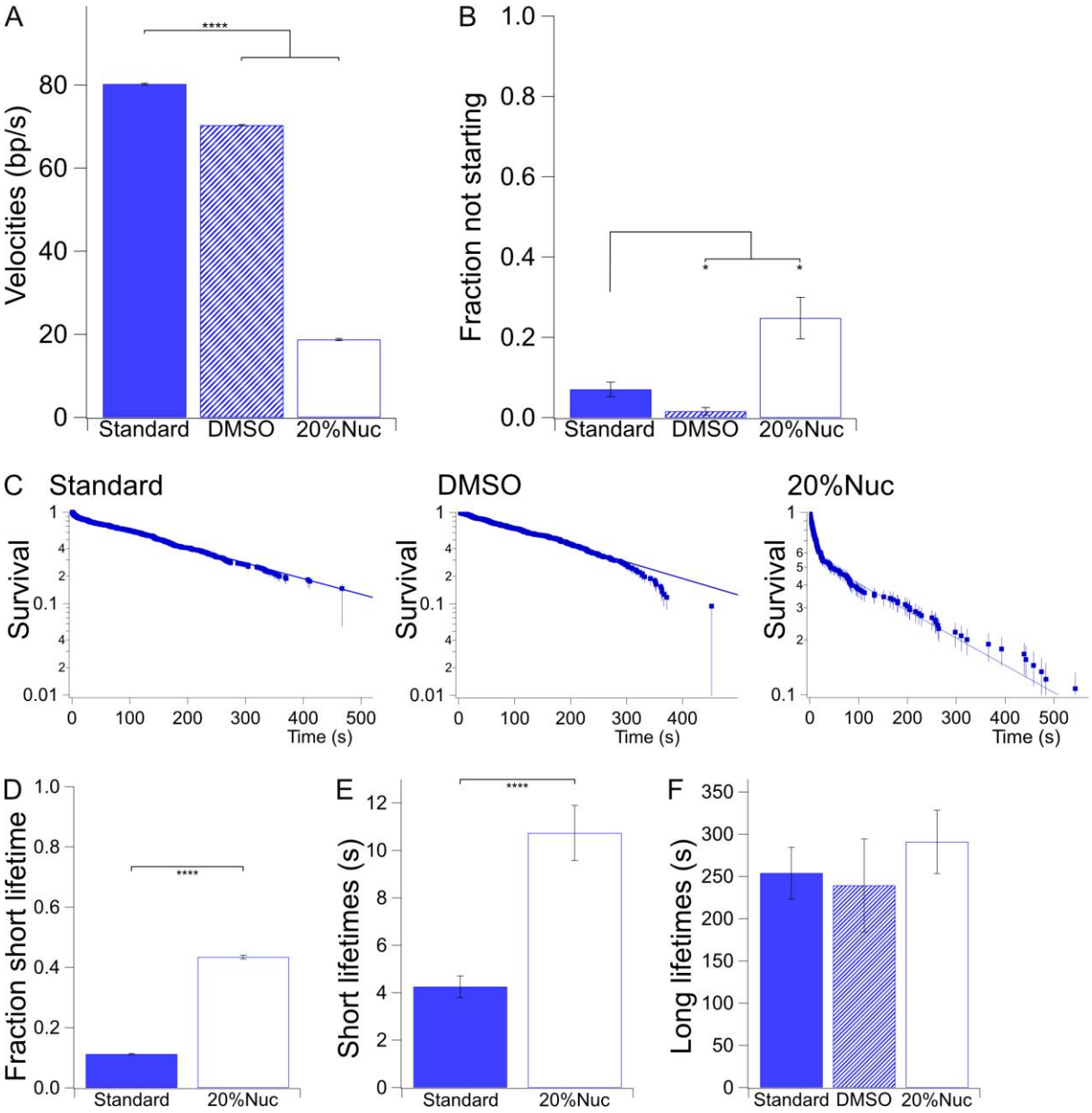


*Figure 4: DMSO increases, and lower nucleotide concentration decreases transcription efficiency. A) Transcription velocities (N = 2281 for standard, N = 4852 for DMSO and p < 0.00001, N = 520 for 20 % Nuc and p < 0.00001, z-tests) of P2 starting polymerases at three different buffer conditions. B) Fraction of not starting polymerases (N = 5 for standard, N = 4 for DMSO p = 0.04, N = 4 for 20 % Nuc p = 0.03, t-tests). C) Survival plots of P2 starting polymerases for three different buffer conditions fitted with a double exponential decay in case of standard buffer conditions (N = 583) or 20 % nucleotide concentration (N = 177) and a single exponential decay in case of DMSO (N = 538). D) Fraction of short lifetime from the double exponential fit from C). (p < 0.00001 z-test). E) Lifetimes of short-binding polymerases from the double exponential fit from C) (p < 0.00001, z-test). F) Lifetimes of long-binding polymerases from the double exponential fit from C) (p = 0.82 for DMSO, p = 0.45 for 20 % Nuc, z-tests). Statistical significance was calculated in comparison to standard buffer conditions for all experiments.*

**A T-rich site causes more frequent transcription termination**

To explore whether transcription elongation is determined by the sequences of the λ-DNA substrate, I analysed kymograms of transcribing T7 polymerases and realised that not all transcription events were unperturbed but instead, polymerases were pausing during transcription initiation (start pause), during transcription (intermediate pause) and before unbinding (end pause, **Figure 5A)**.

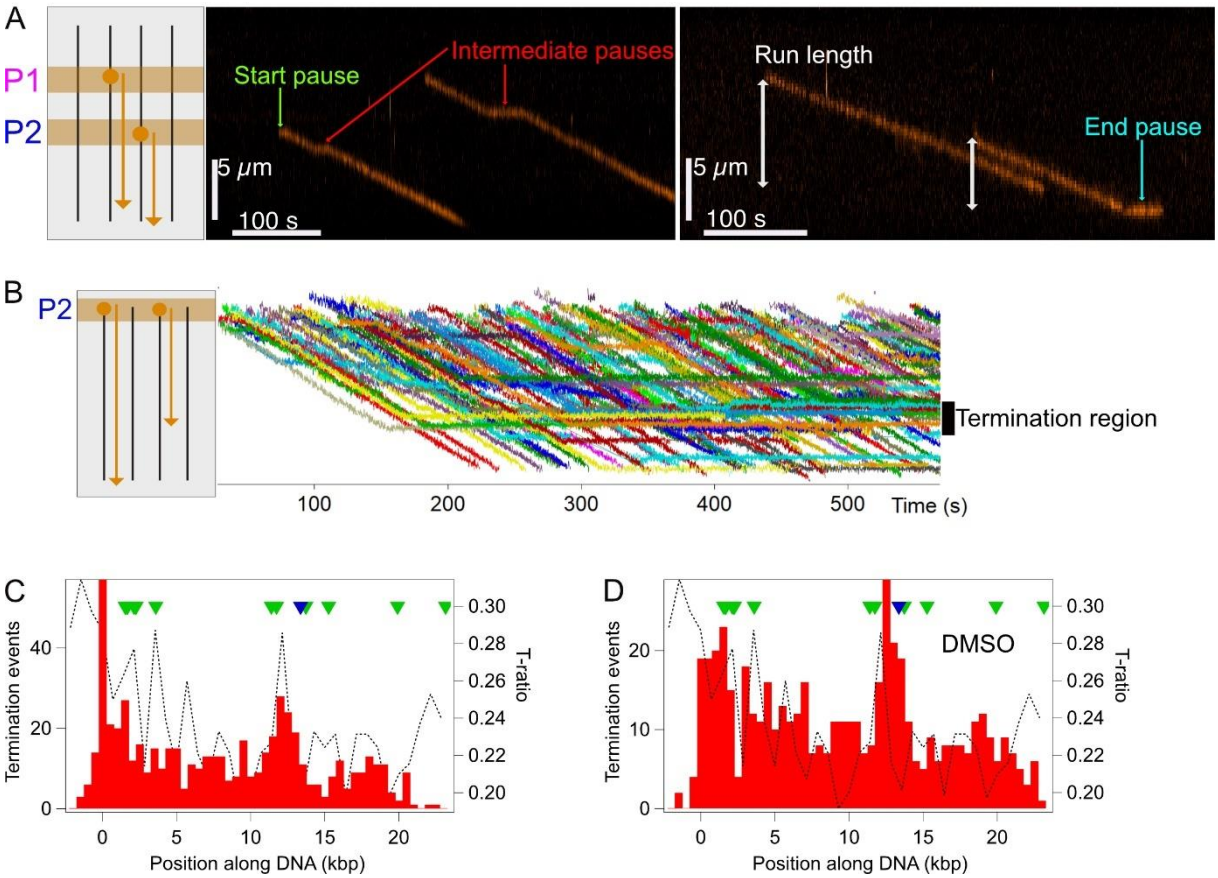



*Figure 5: Polymerase frequently unbinds at a T-rich termination region downstream of the P2 promoter. A) Kymograms of T7 polymerases (orange) starting from P1 or P2 and displaying pauses at the start of transcription (marked with a green arrow), during transcription (marked with red arrows) or at the end of transcription (cyan arrow). White arrows show the run length of transcription events. B) Tracks of all transcription events starting from promoter P2 for a single experiment. Polymerases frequently pause at a T-rich termination region. C) Histogram of run lengths of all*

*transcription events starting from P2 at standard buffer conditions. T-ratio is plotted as a black line on the right axis. Possible termination sites predicted with ARNold are shown as green triangles. Class II terminator sequence shown as a blue triangle. D) As C) but with 10 % DMSO in the reaction buffer.*

When summarising all tracks of starting polymerase of a single experiment **(Figure 5B)**, it became clear that end pauses and transcription termination without a previous pause frequently occurred at a region approximately 12 kbp downstream of P2. This region is characterised by a high T-ratio. I next calculated a histogram of run lengths of transcribing polymerases from P2 to the position of termination **(Figure 5C)** and again observed a peak at this high T-ratio region, meaning that in this region, frequent unbinding occurred. Note that in the presence of 10 % DMSO, the same peak was observed **(Figure 5D)**. This could again be caused by the weaker interaction for uracil-RNA base pairs (bps) (528). I additionally predicted terminator regions using a web tool called ARNold (529) and found two possible hairpin-forming regions directly in front of the mentioned T-rich region and one at the end, which might lead to dissociation of the upstream region in the DNA:RNA hybrid during dissociation on the T-rich region. Additionally, the conserved sequence motif of class II terminator sequences (ATCTGTT (505)) shown as a blue triangle is present at the latter part of the peak, which can further increase the number of early terminations.

I hence conclude that while DNA sequence in the promoter initiation region influences successful transcription initiation, T-ratio within the gene can additionally impact T7 polymerase processivity.

**T7 polymerase can push stalled polymerases off or along the DNA**

As I observed polymerase pauses and early terminations, I wondered what happens to paused polymerases (leading polymerases) if a second polymerase (trailing polymerase) transcribing on the same DNA strand collides with it. *In vivo*, after the production of T7 polymerase by the host polymerase most other genes required for virus assembly are transcribed by the T7 polymerase (491–494), and collision between multiple polymerases transcribing on the same gene might frequently occur. To analyse such collisions *in vitro*, I labelled T7 polymerases with two different fluorescent dyes, LD555 and LD655 and performed transcription experiments using alternating laser excitation with two different laser sources **(Figure 6A)**.

For clarity, the leading polymerase is always marked in pink and the trailing in cyan, independently of the fluorescent label, and the collision event is marked with a yellow star **(Figure 6B)**. For the collision event, average X – and Y – position differences of the two polymerases were calculated **(Figure 6C)**, and collisions for which one or both of the two differences was larger than 1 PSF of the tracked particles (215 nm or 1.21 pxl) were excluded (9 out of 41 events).

In 43.8 % of the cases, the leading polymerase remained at the pause site during the collision **(Figure 6D and Figure 6E top kymogram)**. In these cases it also caused the trailing polymerase to stop, therefore resembling a roadblock. In fewer cases (37.5 %) the leading polymerase was pushed off the DNA **(Figure 6E centre kymogram)**. In all but one of these cases, it also resembled a roadblock, causing the trailing polymerase to stop. In the least number of cases (18.8 %), the trailing polymerase pushed the leading polymerase along the DNA while continuing to transcribe **(Figure 6E bottom kymogram)**.

I conclude that polymerases can clear genes of other stalled polymerases by pushing them off or along the DNA. However, when not pushing the leading polymerase along the DNA, trailing polymerases are almost always stopped themselves, showing that transcription pause sites impair transcription by not only causing the pausing of single polymerases but also leading to collision events detrimental to efficient elongation.

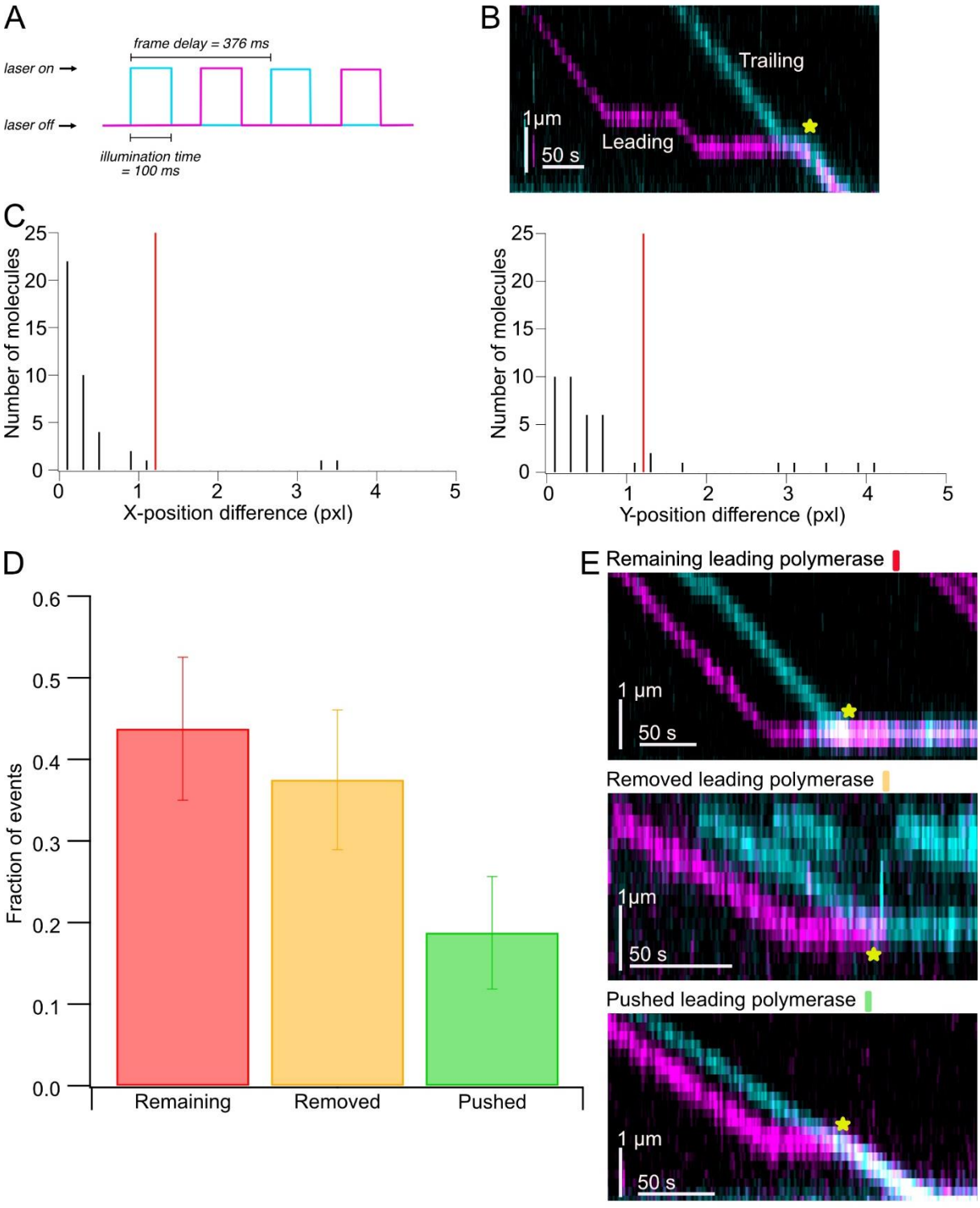



*Figure 6: Trailing polymerases can push leading polymerases off or along the DNA. A) Scheme illustrating laser illumination of 561 nm laser (cyan) and 640 nm laser (pink) during dual-colour experiments. B) Representative kymogram showing a collision event (yellow asterisk) between a paused leading polymerase (pink) and a transcribing trailing polymerase (cyan). C) Average X (left) - and Y (right) - position differences for analysed collision events. If the difference was larger than one PSF of the tracked particles (1.21 pxl or 215 nm) displayed as a red bar, collision events were not counted. D) Fraction of the different events for the trailing polymerase including stopping after collision (red), removed after collision (yellow) or being pushed along the DNA (green). E) Representative kymograms of remaining leading polymerase (top), removed leading polymerase (centre) and pushed leading polymerase (bottom).*

## *1.4 Discussion*

Sequence differences in the initiation region of T7 polymerases can influence transcription efficiency (497, 501, 519). A study using high-throughput sequencing and cDNA amplification identified 5'-TATA**G**GGATAAT-3' as the most and 5'-TATA**G**GGTTCCC-3' as the least efficient initiation sequence (501). Here, I showed that these differences might be caused by differences in the stability of T7 polymerase binding to varying promoter sequences **(Figure 7A)**. I observed similar amounts of binding but a drastic difference in successful transcription initiation on two promoters with identical binding but different transcription initiation sequences. This is caused by a decreased T7 polymerase lifetime on the transcription initiation sequence with a higher T-content, and I assume that this high T-content might lead to more unstable base pairing between uracil-RNA and DNA (528). The sequences of the study above (501) suggest that this might additionally depend on the positioning of the T's and might be more detrimental during the first nucleotide additions when the DNA:RNA hybrid is shorter.

It has been shown that abortive cycling during transcription initiation is not always caused by the polymerase releasing the produced RNA (379, 380) but can also include dissociation of T7 polymerase (382). As RNA was not labelled I can only detect the second case in which both RNA and T7 polymerase unbind from the promoter. RNA release without T7 unbinding shows a different sequence dependence with tight interactions between upstream DNA and T7 polymerase slowing progression into the EC and increasing abortive cycling (524, 525). Therefore unsuccessful initiation depends on two different DNA sequence-dependent mechanisms, abortive RNA release, favoured by tight binding to the promoter sequence stalling transition to the EC (524, 525) and as I show here complete T7 polymerase dissociation depending on different lifetimes on different promoter sequences. I show that lower nucleotide concentration leads to more frequent T7 polymerase promoter unbinding, and in good agreement, it has also been shown to lead to a higher fraction of the second case of abortive cycling in which both RNA and T7 polymerase unbind (382).

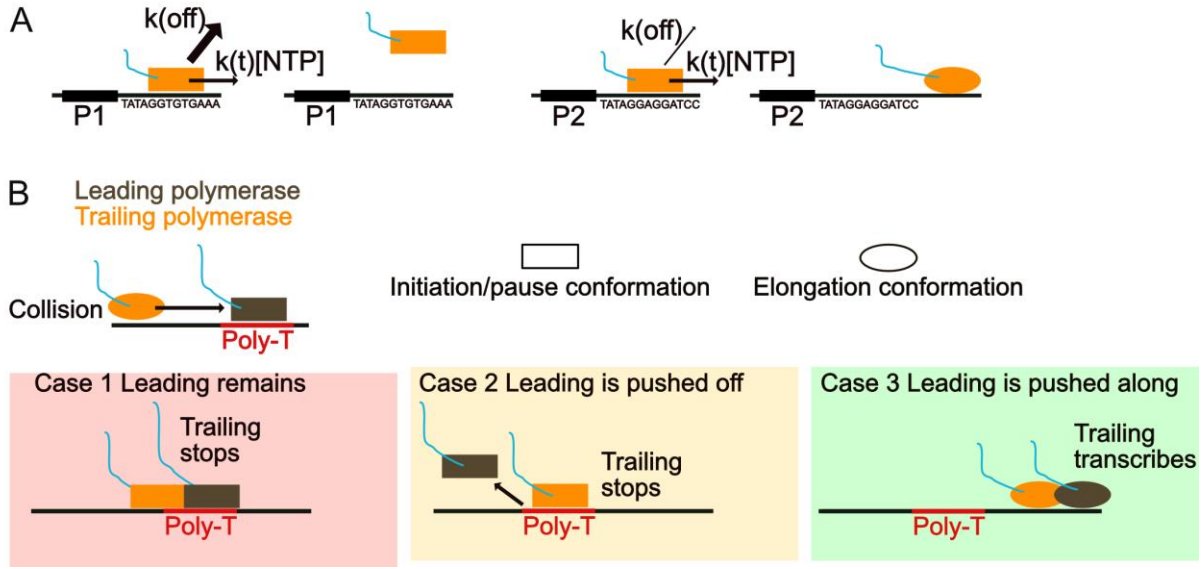



*Figure 7: Model for transcription initiation and transcription collisions caused by T7 pausing. A) Promoter P1 and P2 have the same binding sequence but a different transcription initiation sequence, causing a higher off-rate at P1, where the polymerase unbinds before changing from initiation to elongation conformation. At P2, the polymerase starts transcribing with the nucleotide-dependent rate constant k(t), adopting an elongation conformation. B) Different events during polymerase collisions. If the leading polymerase (brown) remains on the pause site (red box), it resembles a roadblock, causing a stop of the trailing polymerase (orange) at the T-rich pause site*

*(red, Poly-T). Alternatively, the trailing polymerase can push off the leading polymerase leading to a stop of the trailing polymerase in almost all cases (yellow box). In the fewest cases, it was able to continue transcribing by pushing the leading polymerase along the DNA (green box).*

Additionally, I showed that a region with a high T-content also leads to more frequent pausing and transcription termination. For *E. coli* RNA polymerase termination, a forward translocation mechanism has been proposed, which requires melting of the DNA duplex downstream, for example, at destabilising T-rich sequences and upstream RNA hairpin formation pushing the polymerase forward without nucleotide addition, resulting in a shortening of the DNA:RNA hybrid (530, 531). This mechanism can also lead to dissociation of halted T7 polymerase (532). There are two types of termination signals for T7 polymerase. Class I requires a stem-loop-forming sequence and subsequent DNA:RNA hybrid destabilisation by the presence of multiple Us, while class II contains a conserved 7 bp sequence and unbinding is further increased by the destabilisation via downstream Us (505–507, 510), agreeing well with my result of T-rich sequences causing more frequent unbinding. The here used λ-DNA construct does not contain the class I τΦ terminator (507) and no class II VSV or PTH small terminator (505, 509, 533). However, the class II conserved sequence motif (ATCTGTT) (505, 506, 510) is present at the end of the region of observed increased termination. Additionally, there are three hairpin-forming structures predicted by ARNold (529) followed by runs of U within this region, which might lead to the observed increased termination.

I additionally used a dual colour labelling approach to test what happens to paused T7 polymerase complexes when a trailing polymerase collides with them **(Figure 7B)**. While most of the time the leading polymerase resembled a roadblock **(Figure 7B, red box)**, trailing polymerases were also able to clear the DNA of leading polymerases by pushing them off the DNA **(Figure 7B, yellow box)** or along the DNA **(Figure 7B, green box)**, possibly causing a transcription restart. It has been shown that additional transcription reactions on the same DNA template can remove T7 polymerases stalled at DNA damage sites (534). Here, I showed that this can happen by pushing stalled complexes off or along the DNA. Strikingly, while the trailing polymerase was able to push off the leading polymerase, the opposite never occurred, in good agreement with a previous study showing that a trailing polymerase can push off leading polymerases at positions of at least 20 bp from the transcription start sites (TSS) (535). Interestingly, a viral protein called T7 lysozyme binds to the T7 polymerase and inhibits IC to EC transition and also causes increased pausing and termination (513), leading to the assumption that the structural rearrangements happening during IC to EC transition (502) are reversed during pausing or transcription termination (505, 511–513). Paused complexes might, therefore, be in a less stable IC-like conformation (502, 504, 511, 536) and more easily pushed off. During the T7 infection cycle, T7 polymerase and T7 lysozyme are produced by the host polymerase and subsequently, all other genes are transcribed by T7 polymerase (491–494). For efficient transcription of these genes, the virus would require a mechanism to clear its genome of these paused polymerases. Here two mechanisms are revealed, with T7 polymerase being able to both push stalled leading complexes off the DNA or similar to bacterial polymerases (409, 514) push them along the DNA while continuing transcribing themselves.

Additionally, the above results can be applied to optimise promoter and gene sequences for recombinant gene expression and *in vitro* transcription reactions in the biotechnological industry relying on T7 polymerase for example, by the addition of DMSO and usage of promoter sequences with low T-content to enhance transcription initiation and by optimising codons in a way to keep stretches of T to a minimum preventing early termination and pausing.

## II 2. Nucleosome assembly on DNA curtains

**Summary:**
DNA of eukaryotic genomes inside the nucleus is not naked but compacted into nucleosomes, a repetitive structure consisting of ~146 bp of DNA wrapped in 1.75 turns around a histone octamer. Nucleosomes are thereby not equally distributed across the genome with their positioning depending on a variety of different factors including interactions with regulatory proteins, DNA sequence and epigenetic modifications. Positioning of nucleosomes impacts most nuclear processes as regions with high nucleosome occupancy are less accessible for other proteins involved in transcription, DNA replication and DNA repair.

The aim of this project was to assemble nucleosomes on λ-DNA used for a high-throughput single-molecule method called DNA curtains, to be able to study the influence of DNA sequence on nucleosome positioning, while also creating a DNA substrate more closely resembling nuclear DNA for single-molecule *in vitro* studies of other proteins.

To this end, *human* histones were recombinantly expressed, reconstituted to histone octamers, fluorescently labelled and purified. A λ-DNA construct containing nineteen 601 sites (optimised sequences for nucleosome assembly (63)) was created and used to assemble fluorescently labelled histone octamers into nucleosomes. Successful nucleosome assembly was verified by micrococcal nuclease (MNase) digestion. DNA curtain measurements revealed that nucleosomes not only enrich on 601 sites but also form more frequently on GC-rich than on AT-rich regions on λ-DNA.

This enables analysing the influence of transcription factors, polymerases and chromatin remodelling complexes (CRCs) on this positioning in the future and in return also the influence of a more *in vivo*-like DNA construct on the DNA-binding dynamics of these complexes.

**Author contributions:**
Sigrun Jaklin performed histone expressions. Sigrun Jaklin and Jonas Huber performed histone purifications. Jonas Huber fluorescently labelled histone H2B and reconstituted octamers. Jonas Huber prepared a λ-DNA construct containing 19 x 601 sites, assembled nucleosomes and performed MNase assays and single-molecule measurements.

**Abstract:**
DNA curtains are an ideal tool to analyse protein-DNA interactions as hundreds of fluorescently labelled proteins can be visualised on parallelised DNA simultaneously. However, these studies do not accurately represent *in vivo* conditions as eukaryotic DNA inside the nucleus is not naked but forms repetitive structures called nucleosomes, consisting of 146 bp of DNA wrapped around a histone octamer. Nucleosomes therefore regulate DNA accessibility and impact the binding behaviour of other proteins involved in many nuclear processes like transcription, replication and repair. Here a λ-DNA construct containing artificial nucleosome positioning sites was created and used to assemble recombinantly expressed and fluorescently labelled histone octamers into nucleosomes for DNA curtains measurements. Nucleosomes were thereby not only enriched on artificial positioning sites but also preferred GC- over AT-rich regions. This enables the study of both the influence of other DNA-binding proteins on this positioning as well as the influence of a more *in vivo*-like DNA substrate on protein-DNA binding dynamics using DNA curtains.

## *2.1 Introduction*

The chromatin structure formed by nucleosomes impacts the DNA-binding positions and dynamics of regulatory proteins and, thereby, multiple genomic processes such as transcription, DNA replication and DNA damage repair (5, 134, 135, 152, 153, 537, 538). Nucleosomes are formed by ~146 base pairs (bps) of DNA positioned in a left-handed superhelix around a histone octamer, consisting of two copies H2A, H2B forming two dimers and two copies H3 and H4 forming a tetramer (10, 45, 46).

During nucleosome assembly a high amount of bending energy is required to wrap 146 bp of DNA in 1.75 turns around a histone octamer, varying with the underlying DNA sequence, and DNA sequence therefore influences nucleosome positioning (10, 539–541). Nucleosomes have been shown to preferentially form on DNA sequences which display a repetitive reoccurrence of certain dinucleotides at roughly every 10 bp both on genomic DNA (60–62) and on artificial 601 sites (63). Histones form hydrogen bonds via arginine and lysine side chains as well as main chain amide nitrogens with the DNA at roughly every 10 bp when the minor groove is pointing towards the nucleosome (10, 50, 542, 543). On the one hand, it has been shown that arginines preferentially interact with narrow minor grooves found at AT-rich sequences (76), which could explain the periodic reappearance of A/T-containing dinucleotides in 601 sites (63) leading to a higher energetic gain from these interactions and could lead to nucleosomes preferentially forming on AT-rich regions. However, GC-rich sequences are more prone to bend than AT-rich sequences (544–547) explaining the out-of-phase periodic reappearance of G/C-containing dinucleotides in the 601 sites at positions where the major groove points towards the histone octamer reducing the energetic costs of DNA wrapping (50, 62, 63, 544, 545). This raises the question if apart from optimised sequences nucleosomes would preferentially form on AT- or GC-rich regions.

*In vivo,* nucleosomes are not randomly arranged on DNA but have a defined spacing. Nucleosomes form at roughly every 200 bp in *human*s (51, 52) and 165 bp in *yeast* (54, 55) called nucleosome repeat length (NRL). It depends on DNA sequence, as for example in *yeast* AT-rich regions at transcription start sites (TSSs) are mostly depleted from nucleosomes as well as most transcription factor (TF)-binding sites (54, 66, 548). This means that either DNA occupancy by TFs and transcribing polymerases limits nucleosome assembly or causes nucleosome removal or that these DNA sequences are by themselves unfavourable for nucleosomes due to for example AT content. To answer this question, nucleosome occupancy was studied on purified *yeast* genomic DNA (56), which also displayed less nucleosome occupancy on AT-rich sequences showing for example that for five base pair sequences, AAAAA yields the lowest occupancy (56). A high preference for a

10 bp periodicity of dinucleotides was observed, displaying that nucleosomes form on specific sequences even in the absence of TFs and CRCs (56). However, the nucleosome depletion at TSSs was smaller than *in vivo*, especially at positions of highly expressed genes (56) and another study showed the positioning of nucleosomal arrays to mostly not depend on intrinsic DNA sequence preferences (549). In recent years, it has been established that CRCs play an important role in forming arrays of nucleosomes with a conserved spacing around TSSs (59, 143, 144).

Additionally to AT-rich sequences, GC-rich sequences can also cause nucleosome depletion (68), as in mammals, nucleosomes are depleted from GC-rich CpG islands (51, 58, 67), which are bound by different TFs (68–73). This is in contrast to *in vitro* measurements showing increased enrichment on GC-rich compared to AT-rich regions (550) and on simulation data based on the above-mentioned *yeast* genomic DNA study also predicting GC-rich DNA to favour nucleosome assembly (551). What exactly determines nucleosome positioning therefore remains not completely resolved as it depends on many aspects including interactions with TFs (149), action of CRCs (59), intrinsic nucleosome DNA sequence preferences (56) and the action of transcribing polymerases (552).

Different histone variants exist that influence genome accessibility and recruitment of regulatory factors and therefore regulate DNA repair (122), transcription (296) and replication (553). H2A.Z, one of the most common histone H2A variants, plays a role in many nuclear processes (554, 555), including the regulation of transcription (127, 128, 134, 296, 556–559) and DNA accessibility at DNA damage sites (126, 142, 560). H2A/H2B to H2A.Z/H2B dimer exchange frequently occurs around double-strand breaks (DSBs) increasing chromatin accessibility, and the absence of H2A.Z inhibits binding of the repair protein Ku70/80, functioning in non-homologous end joining (126). Additionally, H2A.Z removal is then required for the loading of additional repair proteins (142). H2A.Z seems to play an important role in transcriptional regulation as it is commonly found around promoters (561) and enhancers (127) leading to increased binding of TFs and RNA polymerase (127, 128) and acetylated H2A.Z is commonly found around active genes (296). Positioning of an H2A.Z containing dimer at the (+1) nucleosome (first nucleosome within the gene body), at which the polymerase more frequently pauses than at other nucleosomes within the gene, might lead to increased gene expression and reduced pausing as it is more easily removed from the octamer by the polymerase (134, 554, 559). The *yeast* CRC SWR1 facilitates H2A/H2B to H2A.Z/H2B exchange reactions, by inducing a conformational change in which the DNA is partly unbound from the nucleosome weakening H2A/H2B dimer interactions with the DNA (119, 120). A similar mechanism for the opposite exchange reaction, weakening H2A.Z/H2B dimer contacts with DNA by pushing DNA towards the nucleosome is observed for the CRC INO80 (562), however, INO80's exact roles in H2A.Z positioning *in vivo* are still under debate (563–567). H2A.Z removal within gene regions is caused by transcribing RNA polymerases (568), raising the question if this includes a direct nucleosome disassembly by the force of the motor or requires the recruitment of additional cofactors.

Here, we reconstituted both H2A as well as H2A.Z containing histone octamers and fluorescently labelled them via an H2B cysteine mutant with two different dyes, enabling us to study histone dimer exchange or eviction reactions by different CRCs or polymerases in the future. We assembled fluorescently labelled nucleosomes on modified λ-DNA containing 19 x 601 sites and performed single-molecule measurements on this λ-DNA construct. This showed that nucleosomes formed preferentially on 601 sites while also being more depleted from AT-rich compared to GC-rich regions. This will allow us to study the influence of CRCs like INO80, TFs like CTCF and transcribing polymerases like T7 polymerase on the positioning of assembled nucleosomes in the future as well as the impact of changed DNA accessibility caused by nucleosome occupancy on the DNA-binding positions of these complexes.

**Expression and purification of *human* histones**

Histones were purified according to a previously established protocol (550). *Homo sapiens* (*hs*) histone sequences (generously provided by the group of K.P. Hopfner) for H2B, H2A, H2A.Z, H3 and H4 were inserted into pet21a vectors and cloned into *E. coli Rosetta (DE3) RIL* cells or *E. coli Rosetta (DE3) pLySS* for H4. A threonine119-cysteine (T119-C) mutant of hsH2B was generated. E. *coli Rosetta (DE3) RIL* cells were grown to an optical density (OD) of 0.6 – 0.8 before the expression was induced for 4 h at 37 °C using 0.4 mM IPTG. Cells were centrifugated for 10 min at 3000 rpm and RT. The pellet was solved in 20 ml resuspension buffer (50 mM Tris-HCl, 10 % sucrose and 1 mM benzamidine), frozen in liquid $N_2$ and stored at -80 °C. After thawing, the volume was adjusted to 45 ml, and cells were sonicated for 2 x 45 sec with 40 % output and 40 % duty cycle. After centrifugation at 16 krpm for 20 min at 4 °C pellets were resuspended in TE-100 buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol (β-ME), 1 mM benzamidine) with 1 % Triton. The step was repeated two times in TE-100 buffer without Triton. After final centrifugation, the supernatant was discarded, and pellets were frozen in liquid $N_2$ and stored at -80 °C.

For purification, cells were resuspended in 6.5 ml unfolding buffer (20 mM Tris, 7 M guanidine hydrochloride, 10 mM DTT) and shaken for 1 h at RT. After 20 min centrifugation at 16 krpm, pellets were rinsed with 2.5 ml unfolding buffer, centrifugation was repeated, and supernatants were combined. Cells were dialysed using 6-8 kilodalton (kDa) molecular-weight cut off (MWCO) tubings for 2-3 at room temperature (RT) in 0.5 L urea buffer (10 mM Tris pH 8.0, 7 M urea, 1 mM EDTA pH 8.0, 5 mM β-ME, 100 mM NaCl, deionised by 25 g/L AG 501-X8 resin) followed by dialysis at 4 °C o/n in 1 L urea buffer. For ion-exchange chromatography, 2 x 1 ml Capto Q followed by 2 x 1 ml Capto S ion exchange columns were set up. Columns were equilibrated in a ratio of 90 %/10 % histone buffer 0 (10 mM Tris, 7 M Urea, 1 mM DTT, 1 mM EDTA) / histone buffer 1000 (histone buffer 0 + 1 M NaCl) for H2A, H2A.Z and H2B T119-C and 80 %/20 % histone buffer 0/histone buffer 1000 for H3 and H4 purifications. Protein samples were applied at a flow rate of 0.5 ml/min following a 20 ml wash step in the same buffer ratios as for equilibration. During the last 5 ml of the wash step, the Capto Q column was removed. Histones H2A, H2A.Z, and H2B T119-C were eluted in a 45 ml gradient from 10 % to 55 % histone buffer 1000 followed by 6 ml and 15 ml final elutions at 55 % and 100 % histone buffer 1000 respectively. Histone H3 and H4 were eluted by a 40 ml gradient from 20 % to 60 % histone buffer 1000, followed by 6 ml and 15 ml final elutions at 60 % and 100 % histone buffer 1000. Fractions were analysed using SDS-PAGE, and peak fractions were pooled. Histones were dialysed 4 x against Tris dialysis buffer (10 mM Tris pH 8.0), 2 x with 5 mM β-ME and 2 x without β-ME over 48 h in total using 6-8 kDa MWCO tubings. Samples were frozen in liquid $N_2$ and lyophilised over 48 h before storage at -20 °C.

**Histone octamer reconstitution**

Octamers were reconstituted similarly to a previously established protocol (550). Purified and lyophilised histones were dissolved in unfolding buffer (20 mM Tris–HCl pH 7.5, 7 M Guanidinium-HCl, and 10 mM DTT) with gentle mixing for 2-3 h at RT. Histones were mixed in approximately equimolar ratios (10–15% molar excess of H2A or H2A.Z and H2B relative to H3 and H4), and the concentration was adjusted to 1 mg/ml. For H4, instead of purified H4, commercial H4 (The histone source #HH4) was used. The mixture was then dialysed against refolding buffer (10 mM Tris–HCl (pH 8.0), 2 M NaCl, 1 mM EDTA) using 3.5 kDa MWCO dialysis tubing. The first step (5 h) was carried out using 5 mM β-ME as a reducing agent. For steps 2-4 (8 h or o/n), 5 mM tris(2-carboxyethyl)phosphine (TCEP) was used as a reducing agent. The last dialysis step was carried out for 1 h with 0.2 mM TCEP as a reducing agent. Aggregates were then removed by centrifugation at 4 °C and 3900 rpm for 10 min. Octamers were concentrated to 600 µl by centrifugation at 4 °C

and 3900 rpm using 10 kDa MWCO amicons. A 10x molar excess of Alexa Fluor 568 – maleimide (Thermo Fisher Scientific A20341) or AF647 maleimide (Jena BioScience APC-009-1) over H2B-T119C was added drop by drop, and octamers were labelled at 4 °C o/n. Aggregates were removed by a 10 min centrifugation step at 10 krpm and 4 °C before separating histone octamers from remaining aggregates, tetramers, dimers and free dye on a Superdex 200 size exclusion column. Fractions were analysed on an SDS gel, the octamer peak was pooled, aliquoted and stored at -20°C after the addition of 50 % glycerol.

**Generation of 19 x 601 λ-DNA**

Wild-type λ-DNA was purchased from NEB. For generation of 19 x 601 λ-DNA, a cassette containing a total of nineteen 147 bp 601 sites separated by 53 bp spacer sequences was cloned into the λ-DNA NgoMIV/XbaI site. The ligation product was then packaged, amplified, and functionalised with biotin and digoxigenin according to a previous protocol (229). Instead of size-exclusion chromatography polyethylene glycol (PEG) precipitation was used for subsequent purification to generate higher concentrations of λ-DNA used for nucleosome assembly. To this end, 100 µl of freshly dissolved 30 % PEG-8000 solution was added to 600 µl of the ligation reaction. After addition of 9 µl 1 M $MgCl_2$ solution and o/n incubation at 4 °C, the DNA was centrifugated at 14000 g and 4 °C for 5 min. The supernatant was carefully removed, the position of the DNA pellet was labelled, and the pellet was washed two times with 1 ml 70 % 4 °C cold ethanol solution by centrifugation at 14000 g and 4 °C for 1 min. The supernatant was carefully removed, and the DNA was air-dried at room temperature for 30 – 60 min. Afterwards λ-DNA was resolved in 100 µl TE 150 buffer.

**Nucleosome assembly**

For nucleosome assembly, different ratios of labelled histone octamers (60 : 1 / 150 : 1 / 300 : 1 / 500 : 1 molar excess) and 730 pM 19 x 601 λ-DNA were mixed in TE buffer containing 2 M NaCl and 5 mM β-ME in a volume of 30 µl in custom build dialysis tubings with a 3.5 kDa MWCO. Salt concentration of 300 ml TE 2000 buffer + 5 mM β-ME was reduced during dialysis by adding 2.7 L of TE 50 buffer + 5 mM β-ME at a speed of 3 ml/min using a peristaltic pump. Assembled nucleosomes were stored for up to 1 week at 4 °C before performing DNA curtains or MNase measurements.

**MNase assay**

For MNase assays, nucleosome assemblies were diluted to 60 µl using TE 150 buffer + 5 mM β-ME. Then 7 µl of MNase Master Mix was added (93.5 µl NEB Cut-Smart, 5 µl 1 M $CaCl_2$ and 1.5 µl NEB MNase #M0247S). NEB λ-DNA was used as a negative control. Samples were incubated at 37 °C for 15 min. After the addition of 67 µl MNase stop buffer (20 mM Tris pH 8.0, 50 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid (EGTA), 50 mM EDTA, 1 % SDS) and 3 µl proteinase K (NEB #P8107S) samples were incubated at 50 °C for 1 h. DNAs were purified using a modified protocol of the NEB PCR purification kit (#T1030), adding 6 volumes of ethanol and 2 volumes of NEB DNA cleanup binding buffer to 100 µl of the sample. To 10 µl of eluted sample, 4 µl of 40 % glycerol was added, and samples were analysed on a 1.2 % agarose gel.

**Nucleosome measurements on DNA curtains**

DNA curtain experiments were carried out as described previously (229). Here, 3 µl of nucleosomes assembled on 19 x 601 λ-DNA were diluted to 1 ml in measurement buffer (BSA buffer containing 150 mM NaCl and 1 mM DTT) and added in four 5 min incubation steps to the flow cell. For measurements, an oxygen scavenger system containing glucose-oxidase (Carl Roth), catalase (Sigma Aldrich), and 0.8 % glucose was added as well as 200 nM YOYO-1 (Fisher Scientific #10594233) for DNA visualisation. For nucleosome measurements, videos were recorded with 150 mW (1.05 µW/µm$^2$) 561 nm laser with 100 ms illumination time and 180 ms frame delay.

Enrichment was determined by selecting particles that moved with the DNA upon applied buffer flow using single-tethered (ST) DNA and by creating normalised position data histograms containing 40 bins dividing them into three regions (601 sites, GC-rich and AT-rich) and calculating the enrichment compared to the other two regions. Significance was determined with a two-tailed t-test.

### *2.3 Results*

**Reconstitution of fluorescently labelled H2A and H2A.Z containing histone octamers**

In order to study nucleosomes on DNA curtains, *human* histones were purified using tandem ion exchange chromatography **(Figure 1A-E),** according to a previously established protocol (550). SDS gels of purified histones are displayed in **Figure 1F**. H2A eluted between roughly 150 mM and 330 mM NaCl (A10-D9) and yielded about 32 mg protein. H2B-T119C eluted between 165 and 345 mM NaCl (B1-D12) and also yielded about 32 mg protein. H3 eluted between 240 and 380 mM NaCl (A9-C11) and yielded 10 mg protein. H4 eluted between 400 mM and 450 mM NaCl (D4-D12), giving a comparatively low yield of roughly 0.4 mg. For octamer reconstitution, commercially available H4 was therefore used. H2A.Z eluted between 290 mM and 330 mM NaCl (D1-D9) and gave a yield of about 4 mg protein.

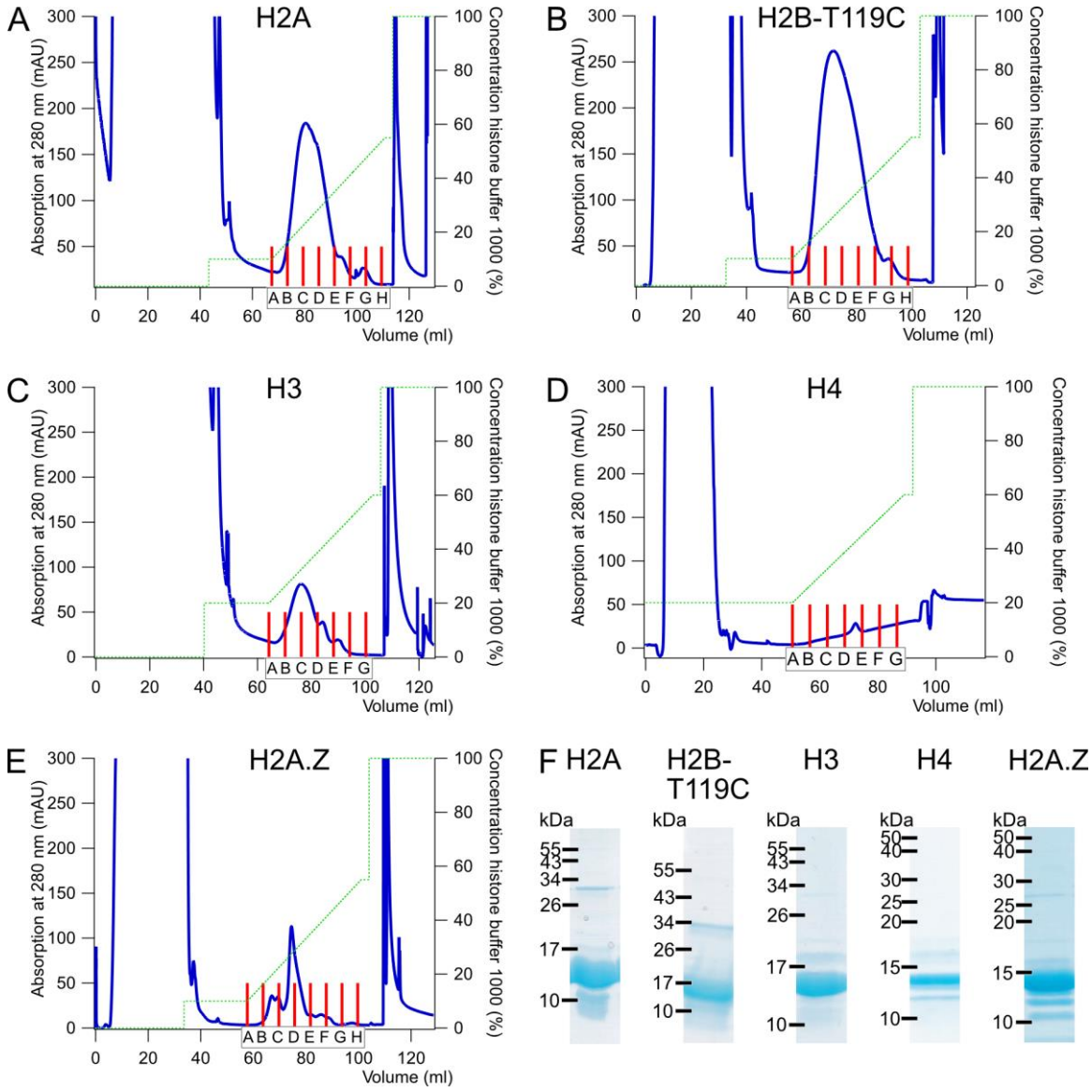


*Figure 1: Ion exchange chromatography of human histones. Human histones were purified via tandem ion exchange chromatography, passing a Capto Q column in front of a Capto S column. The Capto Q column was removed before salt gradient elution. A) H2A (14.1 kDa) ion exchange chromatography fractions showing UV280-absorption in blue (left axis), fractions as red bars, and elution buffer concentration in green (right axis). B) Same as A) but for H2B-T119C (13.9 kDa). C) Same as A) but for H3 (15.4 kDa). D) Same as A) but for H4 (11.4 kDa). E) Same as A) but for H2A.Z (13.6 kDa). F) Representative SDS-gel lanes for purifications of different histones (molecular weights of marker bands displayed by black bars).*

Octamer reconstitution was performed using histones H3, H4 and 10-15 % molar excess of H2B T119-C and either H2A or H2A.Z, minimising H3-H4 tetramer contaminations in the octamer peak during size exclusion chromatography (550). To be able to study the differences between H2A and H2A.Z octamers and visualise histone exchange reactions the cysteine in H2B was labelled with different dyes after octamer assembly **(Figure 2)**. H2A containing octamers were labelled with Alexa Fluor 568 maleimide, which can be visualised by a 561 nm laser and H2A.Z with AF647 maleimide, which can be visualised by a 640 nm laser on the DNA curtains setup. Size exclusion chromatography **(Figure 2A,B)** of reconstituted octamers displayed 4 peaks. The first peak contains aggregated histones, followed by an octamer peak, a dimer peak and a broad peak with two shoulders containing monomers and free dye. Labelling efficiency was 95.5 ± 8.5 % for Alexa Fluor 568-labelled and 67.5 ± 10.4 % for AF647 maleimide labelled H2B-T119C. SDS-PAGE **(Figure 2C)** confirmed the presence of all 4 histones in the purified histone octamers.

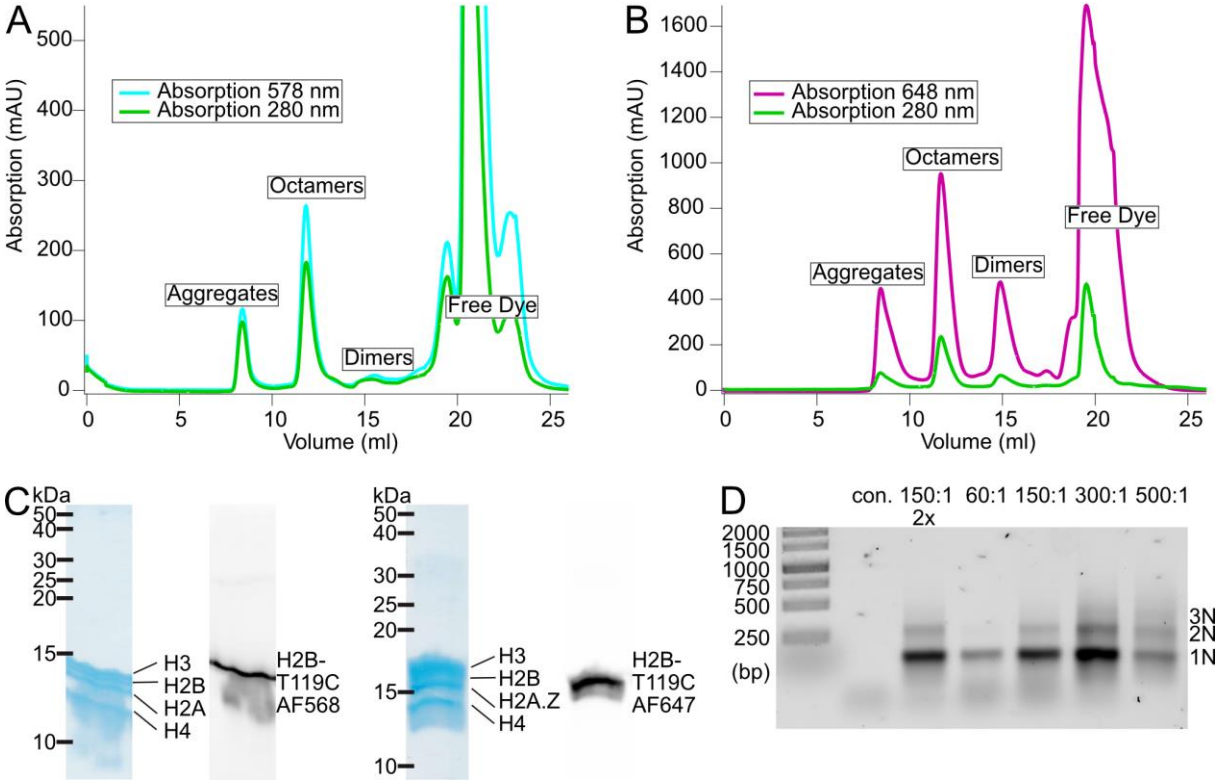



*Figure 2: Purification, labelling and MNase tests of human histone octamers. A) Size exclusion chromatography of H2A containing octamers (UV 280nm in green) labelled with H2B T119-C Alexa Fluor 568 (UV 578 nm in cyan). B) same as A) for H2A.Z containing octamers (UV 280 nm in green) labelled with AF647 maleimide (UV 648 nm in purple). C) SDS-PAGE of reconstituted and labelled histone octamers left: Alexa Fluor 568 labelled H2A containing octamers; right: AF647 labelled H2A.Z containing octamers. (molecular weights of marker bands displayed by black bars). D) Different H2A-octamer to 19 x 601 λ-DNA ratios (60:1 to 500:1) were tested for nucleosome assembly by MNase assay. As control (con.) λ-DNA without octamers was used. In one case (2x),*

*λ-DNA and octamere concentration were doubled (DNA ladder: Thermo Scientific O'GeneRuler 1 kb). 1N, 2N and 3N labels describe the size of DNA being protected from digestion by a single, two or three nucleosomes, respectively.*

**Nucleosomes are enriched on 601 sites and GC-rich regions on λ-DNA**

The next step was to assemble nucleosomes on λ-DNA used for DNA curtain measurements. To achieve this, a DNA fragment containing nineteen 601 sites was cloned into λ-DNA, which was then multiplied by lytic growth and purified by PEG-precipitation and phenol-chloroform extraction. Octamers were mixed with DNA at high salt, and nucleosomes were assembled using salt gradient dialysis. Assembly was verified by MNase digestion, followed by proteinase K treatment. Different molar ratios between 60:1 and 500:1 nucleosome:DNA were used **(Figure 2D)**. As a control (con.), salt gradient dialysis was performed without histone octamers. A ratio of 60:1 (roughly 3:1 molar octamer:601 site ratio) displayed a characteristic band at 147 bp, which is the size of DNA being protected from MNase digestion by a single nucleosome (1N). A ratio between 150:1 and 500:1 (between 8:1 and 26:1 octamer:601 site ratio) displayed additional weaker bands at 294 bp and 441 bp, corresponding to two (2N) or three (3N) nucleosomes next to each other on DNA. Since ratios above 150:1 sometimes led to aggregation, for DNA curtain measurements, a 60:1 or 150:1 ratio was most frequently used.

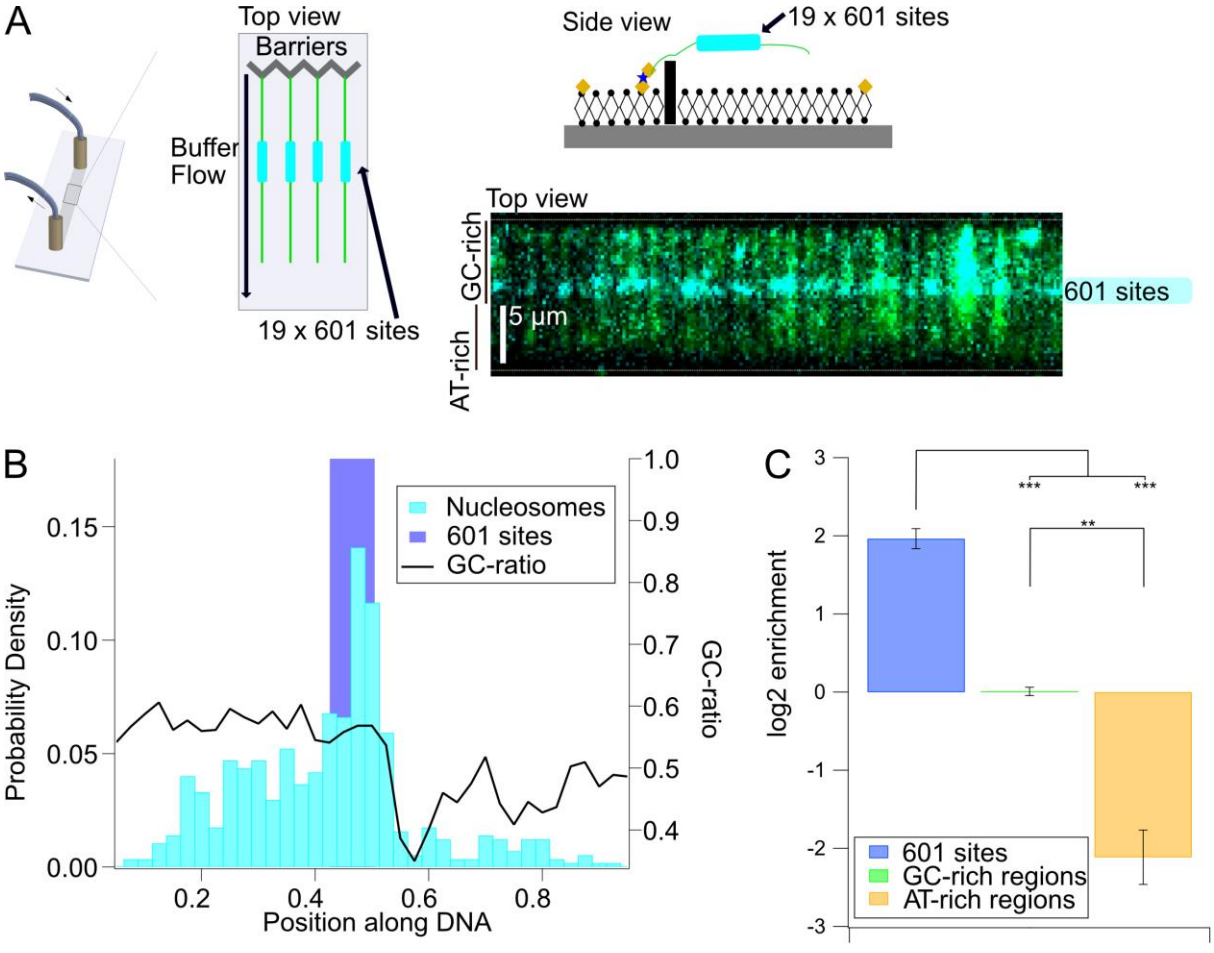



*Figure 3: Nucleosomes enrich on 601 sites and are depleted from AT-rich regions on DNA curtains. A) Schematic representation of a flow cell. Buffer-, DNA- and protein-solutions can be applied to the flow cells by attached inlet- and outlet-tubings. The flow cells contain multiple chromium-nanostructures forming diffusion barriers for λ-DNA. λ-DNA containing 19 x 601 sites is end-labelled with biotin (yellow diamond) bound to a lipid bilayer via biotin-streptavidin (blue star) interaction and transiently stretched by buffer flow. TIRF-image of the DNA curtains displaying most nucleosomes (cyan) are assembled on the 19 x 601 sites on λ-DNA (green), less on GC-rich*

*regions and fewest on AT-rich regions. B) Histogram of nucleosome assembly positions along the DNA (N = 575). GC-ratio is plotted on the right axis. The start and end of the 19 x 601 sites inserted into λ-DNA are marked by a blue square. Most nucleosomes are assembled on the 19 x 601 sites, fewer on GC-rich regions and only few on AT-rich regions. C) log2 enrichment of nucleosome assembly on the 601 sites compared to on GC-rich λ-DNA regions (> 50 % GC excluding 601 sites) and on AT-rich λ-DNA regions. Nucleosomes are significantly more enriched on 601 sites compared to GC-rich regions (p = 0.0001) and AT-rich regions (p = 0.0005) and more depleted from AT-rich compared to GC-rich regions (p = 0.008).*

For DNA-curtain measurements, nucleosomes are assembled on λ-DNA end-labelled with biotin and containing 19 x 601 sites. The biotinylated end is attached via streptavidin to biotinylated lipids and stretched by buffer flow **(Figure 3A)**. A representative image of DNA curtains **(Figure 3A)** displays that most nucleosomes are assembled on the 19 x 601 sites on λ-DNA, fewer on the GC-rich regions at the upper end of the DNA, and almost none on the lower AT-rich end. Nucleosome positions were further analysed in a histogram **(Figure 3B),** again showing enrichment on the 19 x 601 sites. Enrichment was quantified by splitting the DNA into three regions (601 sites/GC-rich/AT-rich) and comparing the enrichment of each to the other two regions. This shows **(Figure 3C)** that nucleosomes show a significantly higher occupancy on 601 sites than on other sites on DNA (roughly 4-fold enrichment), that they display around average occupancy on GC-rich regions and that they are significantly depleted from AT-rich regions **(Figure 3C)**.

### *2.4 Discussion*

Here we successfully created a λ-DNA construct containing fluorescently labelled nucleosomes, which will allow us to study the influence of nucleosome occupancy on transcribing polymerases, genome architecture proteins and CRCs in the feature. Nucleosome assembly was verified using MNase assays showing the assembly of mono- di and trinucleosomes on λ-DNA **(Figure 2D)**.

For DNA curtain measurements as expected a clear enrichment of nucleosomes was detected on the 19 x 601 site, showing that the observed binding events are indeed assembled nucleosomes and not unspecifically bound histones **(Figure 3B,C)**. This could for example be used to study the T7 transcription system developed in **Chapter II 1.** to test if nucleosomes on their own directly cause RNA polymerase pausing. *In vivo,* insertion of a 601 site at the position of the +1 nucleosome causes increased polymerase pausing and reduced RNA production, which might be caused by the tight nucleosome binding to the 601 site but also depended on additional pause factors (391).

In contrast to the 601 sites, we observed almost no nucleosome assembly on the AT-rich segments of our λ-DNA. This is caused by the stiffness of AT-rich regions disfavouring DNA-bending preventing wrapping of DNA around the histone octamer (545), which seems to outweigh favourable energetic interactions of arginine side-chains with narrower minor grooves at AT-rich regions (76). Lower occupancy at AT-rich DNA segments is in good agreement with a previous study on nucleosome assembly on DNA curtains (550) and a result showing that AT-rich segments weaken histone-DNA contacts allowing the binding of other proteins to DNA (64). This can explain the decreased assembly of nucleosomes to AT-rich promoter sequences observed *in vivo* (54, 56, 66) and on purified genomic DNA (56) even so *in vivo* additional factors like the spacing of nucleosomes around TSSs by CRCs (59), polymerases maintaining nucleosome-free regions (NFRs) by promoter-proximal pausing (390) and TFs involved in creating a more open nucleosome structure (149, 569, 570) play a role in generating NFRs around TSSs.

In comparison to the depleted AT-rich regions, GC-rich regions favour nucleosome assembly (51, 551), in good agreement with our data showing more nucleosome assembly at GC-rich than AT-rich

segments. In contrast, *in vivo* CpG islands display decreased nucleosome occupancy (51, 67) depending on GC content (58). CpG islands are frequently non-methylated (571) and could therefore cause increased TF binding (68, 69, 72, 572), which could inhibit nucleosome formation (68). Alternatively, as our GC-rich λ-DNA segments contain an average of 56 % GC content nucleosome enrichment might be observed up to a certain GC richness and disfavoured at even higher GC contents as for example observed for the +1 nucleosome at *mice* promoters *in vivo* with an optimal GC content for nucleosome assembly of 58 % (58).

Most TF-binding sites are positioned at genomic regions of low nucleosome occupancy (573, 574) and most TFs display decreased binding to nucleosome-containing DNA (428) and therefore some TFs require nucleosome remodelling to bind their target sequence (575–577). However, there are also some TFs which bind to histones or to specific DNA sequences within nucleosomes (570, 578) and increase DNA accessibility of previously condensed chromatin (570). Those pioneer transcription factors initially open up chromatin for example by reducing contacts between nucleosomes or weakening histone-DNA contacts of individual nucleosomes (149, 569). However, in some cases, pioneer transcription factors require the action of additional CRCs or histone chaperones for further increasing DNA accessibility and efficient gene expression (149, 150, 579). This makes nucleosome-containing DNA curtains an interesting tool to study different binding positions of TFs depending on DNA sequence, nucleosome occupancy and the TFs ability to bind DNA within nucleosomes as well as the remodelling activities of these pioneer transcription factors.

In summary, we show that *in vitro* nucleosomes are preferentially assembled on 601 sites and also to a lesser extent on GC-rich regions while being mostly excluded from AT-rich regions, allowing us to study the influence of other proteins on this positioning. This could include T7 polymerases, to analyse if they can push nucleosomes along the DNA, similar to what we observed for T7 polymerases pushing CTCF and CTCF-SA complexes off CBSs (229). It could also include CTCF to analyse if nucleosomes interfere with the observed enrichment on the GC-rich CBSs (229) and in return to see if a TF like CTCF influences nucleosome positioning. CBSs *in vivo* are surrounded by nucleosomes with a conserved spacing similar to TSSs (580), raising the question if this is established by CTCF directly or by other interacting factors. It could also include remodelling pioneer transcription factors or CRCs to analyse how they perform their target-site search for nucleosome-bound DNA positions and to study the kinetics of dimer exchange reactions and nucleosome repositioning (119, 120, 140, 149, 570).

## II 3. A conserved ATP- and Scc2/4-dependent activity for cohesin in tethering DNA molecules

Pilar Gutierrez-Escribano, Matthew D. Newton, Aida Llauró, Jonas Huber, Loredana Tanasie, Joseph Davy, Isabel Aly, Ricardo Aramayo, Alex Montoya, Holger Kramer, Johannes Stigler, David S. Rueda, Luis Aragon (230)

**Summary:**

The ring-shaped cohesin complex functions in sister chromatid cohesion preventing early separation of sister chromatids during cell division. It is also required for the formation of topologically associating domains, regulating gene expression and DNA compaction, presumably by capturing a DNA loop inside its ring and enlarging it in a process called loop extrusion. Here we analysed cohesin's cis- and trans-DNA bridging and loop-extrusion capability *in vitro* and their dependence on cohesin's ATPase activity and the cohesin loader by using two single-molecule assays DNA curtains and optical tweezers.

Cohesin's ATPase activity depended on the simultaneous presence of the cohesin loader and DNA, however, ATPase activity was not required to topologically entrap (capture DNA inside its ring) plasmid DNA. DNA curtain measurements allow simultaneous visualisation of fluorescently labelled proteins on parallelly stretched λ-DNA molecules, displaying that cohesin preferentially bound AT-rich DNA sequences and diffused in a salt-dependent but nucleotide-independent manner. Lifetimes displayed a double-exponential decay suggesting two different conformational states of DNA-bound cohesin. Cohesin was able to contract two parallel DNAs, possibly by their entrapment inside its ring, as tethering persisted after washing of non-topologically bound cohesin.

Bridging and tethering behaviour was further analysed using optical tweezer measurements, which are used to apply a force on a DNA molecule tethered between two beads. Force-extension curves of λ-DNA displayed a sawtooth pattern showing rupture events of the DNA compacted by cohesin, which was forming intramolecular DNA bridges in the presence of loader and ATP. Bridges at low salt most often did not withstand extension-relaxation cycles and hence compaction is likely caused by protein-protein interactions instead of ring entrapment. In contrast at high salt, permanent entrapment of two DNA segments took place, resisting multiple DNA extensions at high force. Cohesin was able to capture spatially close-by DNA segments on relaxed DNA simultaneously but was unable to capture a second DNA on relaxed DNA after it had previously been bound to stretched DNA. Cohesin was also unable to compact DNA against an opposing force, suggesting that intramolecular bridges are not formed by motor activity (loop extrusion). However, it can compact single-tethered DNA at low flow rates (low hydrodynamic force) on DNA curtains, most likely a consequence of either loop extrusion or bridging DNA by protein-protein interactions. Performing an assay with two parallel DNAs between four beads displayed that ATPase active cohesin forms bridges in trans, which can move along the DNAs and resist high forces.

This study therefore sheds light on both cellular roles of cohesin. ATPase active cohesin was able to bridge two DNA segments in cis by capturing them simultaneously. As cohesin did not slide off during repeated measurement cycles it captured DNA segments in two different ring compartments or by two interacting cohesin molecules, two potential mechanisms behind TAD formation. It was able to form bridges between two DNA molecules resisting high forces, displaying how cohesin can stably entrap sister chromatids during cell division.

**Author contribution:**

Together with Loredana Tanasie, I performed and analysed ATPase assays of the cohesin tetramer in the presence/absence of DNA and the cohesin loader Scc2/4. I supported Johannes Stigler in measuring cohesin diffusion coefficients on DNA curtains. Together with Loredana Tanasie, I prepared and performed measurements on flipped λ-DNA validating AT-rich binding of cohesin. I was involved in discussions about the data and wrote the ATPase assay method section.

LIFE SCIENCES

# A conserved ATP- and Scc2/4-dependent activity for cohesin in tethering DNA molecules

**Pilar Gutierrez-Escribano[1]*, Matthew D. Newton[2,3]*, Aida Llauró[4], Jonas Huber[5], Loredana Tanasie[5], Joseph Davy[1], Isabel Aly[5], Ricardo Aramayo[6], Alex Montoya[7], Holger Kramer[7], Johannes Stigler[5]†, David S. Rueda[2,3]†, Luis Aragon[1]†**



**Sister chromatid cohesion requires cohesin to act as a protein linker to hold chromatids together. How cohesin tethers chromatids remains poorly understood. We have used optical tweezers to visualize cohesin as it holds DNA molecules. We show that cohesin complexes tether DNAs in the presence of Scc2/Scc4 and ATP demonstrating a conserved activity from yeast to humans. Cohesin forms two classes of tethers: a "permanent bridge" resisting forces over 80 pN and a force-sensitive "reversible bridge." The establishment of bridges requires physical proximity of dsDNA segments and occurs in a single step. "Permanent" cohesin bridges slide when they occur in trans, but cannot be removed when in cis. Therefore, DNAs occupy separate physical compartments in cohesin molecules. We finally demonstrate that cohesin tetramers can compact linear DNA molecules stretched by very low force (below 1 pN), consistent with the possibility that, like condensin, cohesin is also capable of loop extrusion.**

## INTRODUCTION

The establishment of sister chromatid cohesion is essential for accurate chromosome segregation during the mitotic cell cycle. Cohesin is a complex of the SMC (structural maintenance of chromosomes) family originally identified for its role in tethering sister chromatids from S phase until anaphase (*1*, *2*). In addition to its function in sister chromatid cohesion, cohesin modulates the organization of interphase nuclei and mitotic chromosomes (*1*, *3*, *4*). Studies in vertebrates have shown that cohesin complexes maintain contacts between different loci in cis and in this way contribute to the folding of individual chromatids into distinct loops that provide an integral level of genome architecture (*1*, *3*, *4*). The current model for how SMC complexes, including cohesin, might form DNA loops involves the capture and bending of DNA segments followed by progressive enlargement of these to form loops (*5*, *6*); this activity has been termed "loop extrusion." Evidence for this model has been obtained from in vitro analysis of purified yeast condensin (*7*). Cohesin's most prominent function is the tethering of sister chromatids, which is expected to involve an ability to bridge two DNA molecules in trans. Unlike condensin, cohesin has not yet been demonstrated to extrude loops in vitro. A potential activity in loop extrusion has been suggested for cohesin because of its involvement in the maintenance of cis looping and as a potential linear tracking mechanism that could explain the preferential use of convergent CTCF DNA motifs at TAD borders during genome folding (*8*, *9*). However, it is currently not clear how loop extrusion could explain the well-established role of cohesin in sister chromatid cohesion.

Mechanistically, we only have a vague idea of how cohesin might generate intermolecular tethers while mediating sister chromatid cohesion. Two main models have been proposed to explain cohesin function in sister chromatid cohesion: the "ring" or "embrace" model (*10*, *11*), in which a single cohesin ring entraps both sister DNA molecules (*10*), and the "handcuff model," where sister chromatid cohesion is mediated by the entrapment of sister DNAs in different cohesin complexes and a subsequent cohesin-cohesin interaction (*1*, *12*, *13*). The capture of the pair of double-stranded DNA (dsDNA) molecules during the establishment of sister chromatid cohesion by a single cohesin molecule in the "embrace model" has been proposed to occur by either (i) passage of the replisomes through the ring lumen of a DNA-bound cohesin or (ii) when a DNA-bound cohesin captures a single-stranded DNA (ssDNA) at the fork, which is then converted into dsDNA by DNA synthesis (*14*). Although cohesin complexes have been purified from fission yeast (*15*), frogs (*16*), and human cells (*17*), single-molecule analyses of DNA bridging activities have not been reported. Purified cohesin complexes have been shown to exhibit DNA binding activity in a salt-resistant manner (*18*) and to rapidly diffuse on DNA; however, these were shown to be independent of adenosine triphosphate (ATP) (*15*–*17*), suggesting that they are not at the core of its ATP-dependent activity.

Single-molecule studies of purified yeast condensin have shown that this SMC complex compacts DNA molecules on magnetic tweezers (*19*), translocates along linear DNA molecules in an ATP-dependent manner (*20*), and forms DNA loop–like structures on surface-tethered, flow-stretched DNA (*7*). Furthermore, while purified condensin exhibits robust ATPase activity in the presence of DNA (*19*), purified yeast cohesin is a poor ATPase on its own (*21*, *22*). Recent work has shown that the Scc2-Scc4 loader complex greatly stimulates cohesin's ATPase activity (*21*, *22*). On the basis of these findings, we sought to investigate activities of budding yeast cohesin in the presence of the Scc2-Scc4 loader complex using the

[1]Cell Cycle Group, Medical Research Council London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK. [2]Single Molecule Imaging Group, Medical Research Council London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK. [3]Molecular Virology, Department of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK. [4]LUMICKS, De Boelelaan 1085, 1081 HV, Amsterdam, Netherlands. [5]Gene Center, Ludwig-Maximilians-University, 81377 Munich, Germany. [6]Microscopy Facility, Medical Research Council London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK. [7]Biological Mass Spectrometry and Proteomics Facility, Medical Research Council London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK.
*These authors contributed equally to this work
†Corresponding author Email: stigler@genzentrum.lmu.de (J.S.); david.rueda@imperial.ac.uk (D.S.R.); luis.aragon@csc.mrc.ac.uk (L.A.)

following two complementary single-molecule approaches: DNA curtains and optical tweezers.

## RESULTS

### Purification of cohesin complexes

To investigate activities of yeast cohesin using single-molecule assays, we first purified budding yeast cohesin tetramers, containing Smc1, Smc3, Scc1/Mcd1 (thereafter referred to as Scc1), and Scc3, from exponentially growing yeast cultures (Fig. 1A). Cohesin subunits were overexpressed in high-copy plasmids using galactose (*GAL*)–inducible promoters. Purified material was obtained via affinity chromatography, using a triple-StrepII tag fused to the Smc1 subunit, followed by passage through a HiTrap Heparin HP column (Fig. 1A and table S1). Analysis of purified complexes by negative-stain electron microscopy confirmed the presence of rod-shaped cohesin holocomplexes, the majority in a folded conformation (Fig. 1B) (*23*). The Scc2-Scc4 complex was also purified from budding yeast (Fig. 1C) using a similar strategy and showed DNA binding activity as expected (fig. S1A) (*21*, *22*). Purified cohesin also bound plasmid DNA in a salt-resistant manner (fig. S1B),

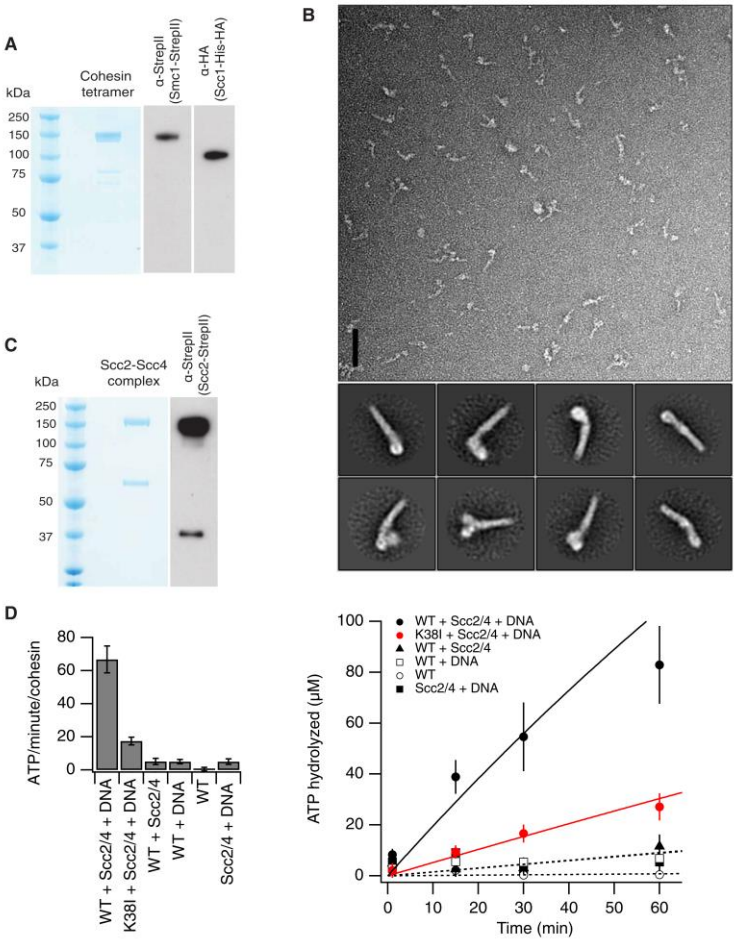



**Fig. 1. Biochemical purification of yeast cohesin.** (**A**) Purified cohesin tetramer containing Smc1, Smc3, Scc1, and Scc3 was analyzed by SDS–polyacrylamide gel electrophoresis (PAGE) followed by Coomassie blue staining. Western blot analysis shows the mobility of Smc1 and Scc1. (**B**) Top panel: Representative micrograph of a BS3-crosslinked cohesin sample observed in negative stain EM. Scale bar, 50 nm. Bottom panel: Class averages obtained with RELION. A set of the best ~5000 particles was used for this classification. The size of the circular mask is 450 Å. (**C**) Coomassie blue staining of the purified Scc2-Scc4 complex. (**D**) ATP hydrolysis by yeast cohesin and cohesin ATPase mutant Smc3-K38I with or without the Scc2-Scc4 complex.

and the bound plasmid was released by DNA cleavage with restriction enzymes (fig. S1C). This is consistent with the topological binding mode proposed for this complex (*18*, *22*). However, in our hands, this activity was not strictly dependent on ATP and was not stimulated by Scc2-Scc4 (fig. S1B), in contrast to what has been reported recently (*18*, *22*). Last, we confirmed that our purified Scc2-Scc4 complex was able to stimulate cohesin ATPase activity (Fig. 1D) (*21*, *22*).

### Activity of yeast cohesin tetramers on double-tethered DNA curtains

Next, we sought to test whether budding yeast cohesin exhibited the behavior described for cohesin from other organisms on DNA curtains (*15*–*17*). λ-DNA molecules (48.5 kb) were anchored to a lipid bilayer in a flow cell surface and aligned into double-tethered DNA curtains using nanofabricated barriers (Fig. 2A) (*15*). Quantum dots (Qdots) conjugated to antibodies against the hemagglutinin tag (HA3) fused to the C-terminal region of the Scc1 kleisin subunit were used to visualize the complexes (Fig. 2B). On flowing the labeled cohesin complex over the DNA curtains, binding was observed at low ionic strength (Fig. 2A). The chamber was flushed with a high ionic strength buffer to remove nontopologically bound complexes (Fig. 2A). While a large fraction of cohesin complexes dissociated, we observed diffusion along the DNA (Fig. 2B). The binding preference of cohesin to more A/T-rich regions reported earlier (*15*) was also observed (Fig. 2, C to E). The diffusion coefficients correlated with the ionic strength of the buffer (fig. S2F). The survival probabilities of cohesin were not affected by the addition of ATP, or the ATP analogs adenosine 5′-diphosphate (ADP) and ATPγS (Fig. 2C). We found that the presence of Scc2-Scc4 enhanced the ability of cohesin to stay bound on the DNA (Fig. 2D); however, the presence of nucleotides did not alter cohesin stabilities (Fig. 2D). Therefore, these results are consistent with the activities observed for cohesin from other organisms (*15*, *17*) and show that budding yeast cohesin undergoes rapid diffusion on DNA curtains in an ATP-independent manner.

In our DNA curtain experiments, we made an observation not reported in earlier studies (*15*, *17*). Cohesin signals were often observed bound between what appeared to be two fused DNAs (Fig. 2E). The pairing events formed under low-salt conditions in the presence of ATP (Fig. 2F and movies S1 and S2), but they persisted when the chamber was flushed with a high ionic strength buffer, raising the possibility that topologically bound complexes mediated these events.

### Cohesin can form two classes of DNA bridges in an ATP- and Scc2-Scc4–dependent manner

To further explore our observation that cohesin tetramers paired λ-DNA molecules on the DNA curtains, we decided to use a dual-trap optical tweezer with confocal fluorescence microscopy capabilities. A similar approach has been previously used in the study of protein-DNA interactions (*24*). Briefly, we tether a λ-DNA molecule with biotinylated ends to two optically trapped streptavidin-coated polystyrene beads, enabling us to accurately apply and measure forces on the captured DNA molecule. We performed our experiments in multichannel laminar flow cells where we had the possibility to move the tethered DNA to different flow lanes containing different protein complexes and buffers. In addition, we were able to image the tethered DNAs using confocal fluorescence microscopy. Overall, the approach allows increased experimental control over DNA curtains. Proteins can be added, removed, or incubated under different salt conditions sequentially, and the physical effect of their activities can be measured accurately on a single DNA molecule.

To test for the formation of intramolecular cohesin bridges in cis, we adapted a previously published protocol that measures protein-mediated DNA bridging (Fig. 3A) (*25*, *26*). First, we captured a single λ-DNA molecule and generated a force-extension (FE) curve in the absence of protein by extending the molecule slightly beyond its contour length (~16 μm). We then moved the DNA to a channel containing 1 nM cohesin, 2.5 nM Scc2-Scc4 complex, and 1 mM ATP in 50 mM NaCl and incubated for 30 s in a relaxed conformation (~3 μm between beads). Following incubation, the relaxed DNA was then moved to a channel without protein but containing 1 mM ATP in 125 mM NaCl. Reextending the DNA in the buffer-only channel yielded FE curves with sawtooth features at extensions shorter than the contour length (Fig. 3B, Cohesin + Scc2/4). This is characteristic of intramolecular bridge rupture events (*25*, *26*) (Fig. 3A, right) and shows that cohesin can tether the DNA in cis forming a protein-mediated bridge between different segments of the molecule, thus creating an intramolecular loop. When we repeated this protocol in the presence of 1 nM cohesin and no Scc2-Scc4 (Fig. 3B, Cohesin), or 2.5 nM Scc2-Scc4 and no cohesin (Fig. 3B, Scc2/4), FE curves identical to those of the initial naked DNA were observed, demonstrating that no protein-mediated bridges were formed (Fig. 3A, left). Similarly, incubating 1 nM cohesin and 2.5 nM Scc2-Scc4 complex in the absence of ATP, or with the ATP analogs ADP or ATPγS, yielded FE curves identical to those of naked DNA (fig. S3A). To confirm the requirement of ATP, we repeated the protocol in the presence of 1 nM cohesin ATPase mutant (K38I) (fig. S4) and 2.5 nM Scc2-Scc4 (Fig. 3B, CohesinK38I + Scc2/4). FE curves identical to those of the naked DNA were observed (Fig. 3B, CohesinK38I + Scc2/4). Therefore, the DNA bridging activity requires ATP and depends on the Scc2-Scc4 loader complex.

Next, we tested the effect of ionic strength on cohesin bridging (Fig. 3C). Cohesin bridges were observed at all salt concentrations tested (Fig. 3C). The length of DNA extension released during the rupture of a DNA bridge can be directly related to the loop size encompassed by the bridge. We analyzed the sizes of the DNA bridges from the FE curves at 125 mM salt (fig. S3B) and found that the distribution of loop sizes is exponential with a characteristic size of ~900 base pairs (bp), consistent with a model of random bridge formation (*5*, *6*). Most of the small sawtooth peaks observed at low forces and extensions disappear under high-salt conditions, while the overall contour length of the DNA remained reduced (Fig. 3C). We also recorded FE curves when we relaxed tethers (Fig. 3, B and C, reverse arrows) after the extensions are done in the buffer channel, therefore in the absence of protein (Fig. 3, B and C, forward arrows). These showed that compaction due to DNA bridges formed at low-salt concentrations was lost after extension (Fig. 3C, reverse arrows; 50 mM NaCl) with force. However, relaxation of tethers with DNA bridges formed at high-salt concentrations showed compaction events that had resisted after extension (Fig. 3C, reverse arrows; 300 and 500 mM NaCl). These results show two distinct types of cohesin bridging events: (i) one predominantly occurring at low salt that is characterized by frequent interactions that are "reversible" and can be disrupted by moderate force (5 to 40 pN) and (ii) a second "permanent" bridge class that resists higher ionic strength conditions and full physical stretching of the DNA molecule. Both classes of DNA bridges were not observed when an ATPase mutant cohesin

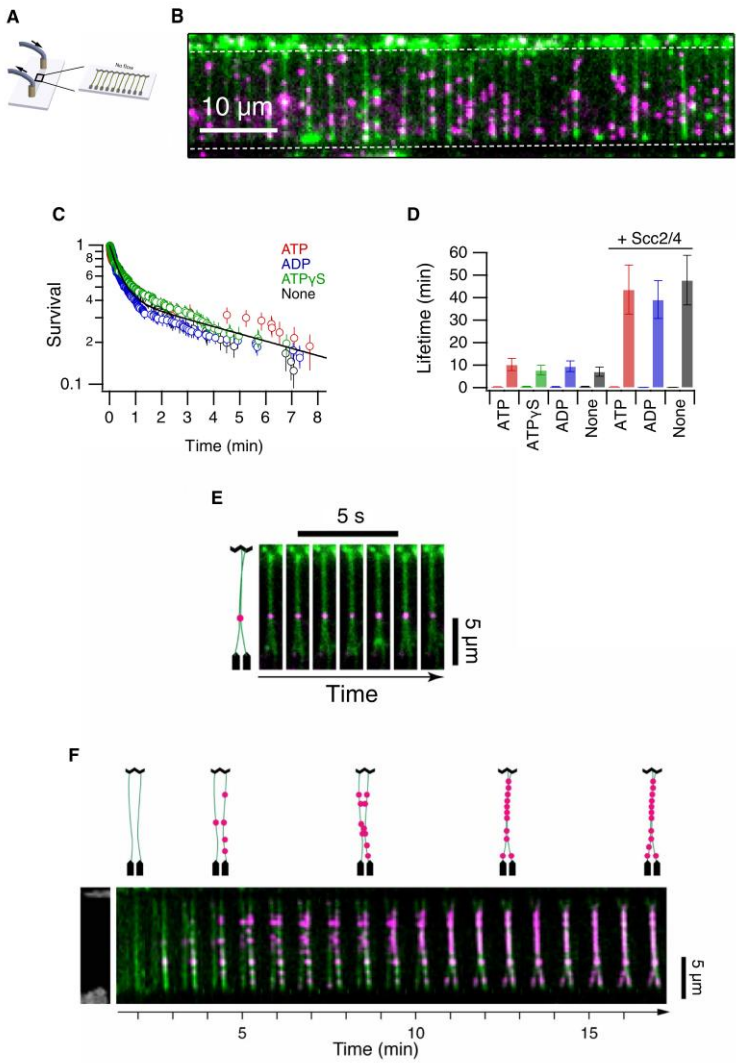



**Fig. 2. Analysis of yeast cohesin on DNA curtains.** (**A**) Schematic representation of double-tethered DNA curtains used in the study. (**B**) Image of cohesin tagged with quantum dots (magenta) bound to λ-DNA stained with YOYO-1 (green). Scale bar, 10 μm. (**C**) Survival probability plots of cohesin in the presence of ATP, ADP, ATPγS, or no nucleotide. (**D**) Lifetimes of cohesin (fast phase and slow phase) in the presence or absence of Scc2-Scc4 and different ATP analogs. Error bars are 68% confidence intervals from bootstrapping. (**E**) Image of a pair of double-tethered DNA curtains bound by cohesin. DNA molecules are in green, and cohesin is in magenta. Diagrammatic representation is shown (left). (**F**) Time-lapse images of a pair of double-tethered DNA curtains bound by cohesin as they are tethered. DNA molecules are in green, and cohesin is in magenta. Diagrammatic representation is shown (top). Pairing events were observed frequently in the DNA curtains. An average of 5 to 10 events per DNA curtain was detected.

complex (SMC3-K38I) was used (Fig. 3B, CohesinK38I + Scc2/4), confirming that the ATPase activity of the complex is a requirement for both types of bridges. Next, we tested whether permanent bridges could resist repeated extensions. We performed two cycles of bead extension and relaxation and confirmed the persistence of the permanent cohesin bridge using FE curves (fig. S5). We conclude that permanent cohesin bridges resist high stretching forces and that the complexes mediating the tethers cannot be displaced from the DNA

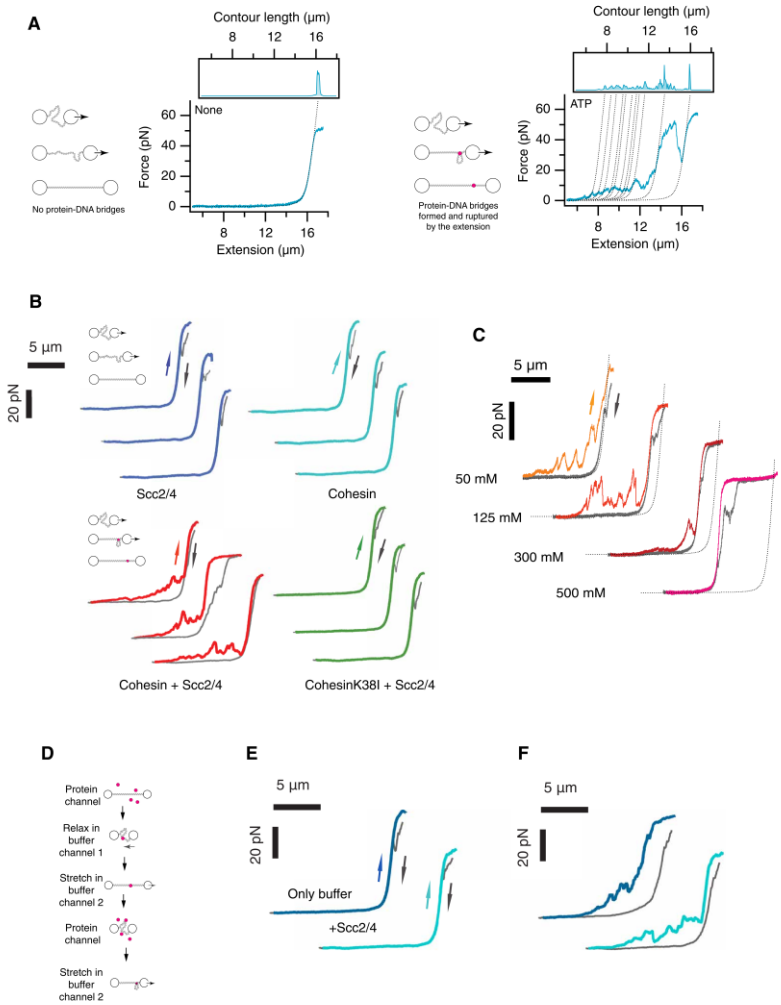



**Fig. 3. Cohesin bridges DNA in an ATP- and Scc2-Scc4–dependent manner.** (**A**) Schematic representation of FE curve for λ-DNA exhibiting the presence (right diagram and graph) and absence (left diagram and graph) of protein DNA bridges. Dotted line is fit to worm-like chain for naked DNA. (**B**) FE curves for λ-DNAs preincubated with 1 nM cohesin and 2.5 nM Scc2-Scc4 complex and 1 mM ATP (Cohesin + Scc2/4), 1 nM cohesin and 1 mM ATP (Cohesin), 2.5 nM Scc2-Scc4 complex and 1 mM ATP (Scc2/4), or 1 nM cohesin ATPase mutant and 2.5 nM Scc2-Scc4 complex and 1 mM ATP (CohesinK38I + Scc2/4). Schematic diagram of the experimental design. After capturing a single DNA molecule between two optically trapped beads, DNA was incubated in the presence of protein in a relaxed conformation (3-μm bead distance) for 30 s in 50 mM NaCl and then moved to a buffer channel with 125 mM NaCl for extension and measurements. Only incubation with 1 nM cohesin and 2.5 nM Scc2-Scc4 complex and 1 mM ATP (Cohesin + Scc2/4) showed DNA bridging rupture events. (**C**) FE curves in the presence of increasing ionic strength. High salt favors topologically constrained and permanent DNA bridges. (**D**) Schematic representation of the experimental design to test cohesin second DNA capture. After capture of λ-DNA between the two optically trapped beads, DNA is extended and incubated for 30 s in the protein channel. DNA is moved to a buffer channel and then relaxed (3-μm bead distance) and incubated for 30 s before reextension to test for DNA bridges (E). The extended DNA is then incubated in a relaxed position in the protein channel and then moved to buffer channel and extended to confirm that bridges can be formed when protein is loaded while DNA is relaxed (F). (**E**) λ-DNA incubated with 1 nM cohesin, 2.5 nM Scc2-Scc4 complex, and 1 mM ATP in an extended conformation and then moved to a buffer channel (125 mM NaCl) in the presence of 1 mM ATP (buffer only, dark blue) or 2.5 nM Scc2-Scc4 complex and 1 mM ATP (+Scc2/4, light blue). DNAs were reextended, and the FE curves were recorded. (**F**) The λ-DNA molecules in (E) were incubated in a relaxed position (3-μm bead distance) in the presence of 1 nM cohesin, 2.5 nM Scc2-Scc4 complex, and 1 mM ATP DNAs. DNAs were moved to a buffer-only channel (125 mM NaCl containing 1 mM ATP) and reextended. FE curves show the presence of DNA bridging rupture events.

molecules. This explains the repeated detection of the same bridge on FE curves during the cycle of bead extension and relaxation (fig. S5).

### The tethering of two dsDNA segments by cohesin occurs in a single-capture step

Recent studies using purified cohesin from *Schizosaccharomyces pombe* have shown that cohesin can capture a second DNA, but only if single stranded (*14*). This led the authors to conclude that cohesin is not capable of trapping to dsDNAs in vitro (*14*). Moreover, it was suggested that this activity is likely to occur at replication forks, where cohesin bound to a dsDNA molecule is exposed to nascent ssDNA (*14*). The second capture of the single-stranded molecule was dependent on the presence of cohesin loader and ATP (*14*). Our results seem to contradict this because we show that cohesin purified from *Saccharomyces cerevisiae* is fully able to trap two dsDNA molecules (Fig. 3, B and C). Next, we decided to investigate whether capture of the two molecules is sequential or simultaneous. In our original tethering assay, we could not differentiate whether the two dsDNAs are captured sequentially or in a single step, as we had incubated the DNA in a relaxed position (with the two DNA segments in proximity). To distinguish whether one or two events were involved in the formation of the cohesin tethers observed, we sought to test whether cohesin could capture a second DNA after initial loading. To this aim, we captured a single λ-DNA molecule and generated an FE curve. We maintained the DNA in an extended position (~15 μm between beads) using a pulling force of 5 pN (Fig. 3D) and loaded cohesin by moving the DNA to a channel containing 1 nM cohesin, 2.5 nM Scc2-Scc4 complex, and 1 mM ATP in 50 mM NaCl. We incubated the DNA for 30 s (Fig. 3D) before moving it to a different channel containing 1 mM ATP in 125 mM NaCl. We then relaxed the DNA conformation (~3 μm between beads) to allow DNA segments to come into proximity (Fig. 3D) and incubated in the relaxed conformation for an additional 30 s. The FE curve obtained after reextension of the DNA was identical to the initial naked DNA profile (Fig. 3E, Only buffer, and fig. S6). We obtained a similar result when we included 2.5 nM Scc2-Scc4 complex and 1 mM ATP in the channel where we relaxed the DNA (Fig. 3E, +Scc2/4, and fig. S6). These results show that loaded cohesin is unable to capture a second DNA segment. To confirm that DNA bridges could be formed in the same DNA in one step, we relaxed the molecules used in the experiments and incubated them for 30 s in a channel containing 1 nM cohesin, 2.5 nM Scc2-Scc4 complex, and 1 mM ATP. When molecules were reextended, the resulting FE curves confirmed the formation of DNA bridges (Fig. 3F and fig. S6). In addition, we confirmed that cohesin complexes can bind to extended DNAs using a published DNA friction protocol (fig. S7) (*27*). Therefore, our results are consistent with a previous report (*14*), showing that cohesin bound to DNA cannot undergo a second capture event involving a dsDNA molecule, but demonstrate that cohesin is able to capture two dsDNAs simultaneously. A previous study could not evaluate the possibility that cohesin could capture two dsDNAs simultaneously, thus reaching an erroneous conclusion (*14*). We conclude that cohesin establishes bridges between two dsDNA in a single step, or two kinetically very close steps, which requires physical proximity of the DNA segments.

### Cohesin can slide while tethering two DNAs intermolecularly

Next, we investigated whether cohesin can form intermolecular bridges. We developed an intermolecular bridging assay, where two dsDNA molecules are tethered in parallel between the pair of beads, and tested the ability of cohesin to form bridges between these two molecules (Fig. 4A). DNA molecules were visualized with SYTOX Orange. After confirming the presence of two DNA molecules tethered in parallel between the beads (Fig. 4B, Naked), the DNA was incubated in a relaxed state to bring the DNAs into proximity (~3-μm bead distance) in the presence of 1 nM cohesin and 2.5 nM Scc2-Scc4 and 1 mM ATP for 30 s. The DNAs were moved to a buffer-only channel (300 mM NaCl plus 1 mM ATP). Strikingly, clear bridging was observed between the two molecules on reextension (Fig. 4B, Cohesin + Scc2/4). DNA bridges did not form in the absence of ATP (Fig. 3B, no ATP) or when we used cohesin ATPase mutant complex (Fig. 4B, K38I + Scc2/4), confirming that cohesin's ATPase activity is required. Bridge formation in this assay was very efficient; of 10 molecules tested, 8 showed intermolecular bridges (Fig. 4B, Cohesin + Scc2/4) and 2 showed intramolecular bridging on the two individual DNAs. Intermolecular bridges always appeared to be near the midpoint of the DNA (Fig. 4B, Cohesin + Scc2/4). Potential reasons to explain this include the fact that the central region of λ-DNA molecules is rich in A/T content where cohesin might bind preferentially. Alternatively, cohesin might be able to slide on the DNA while maintaining tethers and therefore likely to move to the center regions as the molecules are extended. To further characterize this, we used a quadruple-trap optical tweezer setup, which allows the independent manipulation of the two DNA molecules (*27*).

We first captured two single λ-DNA molecules using a pair of traps for each (DNA1 between traps 1 and 2 and DNA2 between traps 3 and 4) in a parallel conformation (fig. S8). Both DNA molecules were stretched close to their contour lengths (~16 μm). We then manipulated DNA2 using beads 3 and 4 and moved it upward (in the *z* direction) before rotating it 90° and moving it into a crossed conformation directly above DNA1 (fig. S8). We then lowered DNA2 to its original *z* position and relaxed it to ensure physical contact between the two DNA molecules at the junction point (fig. S8). We then moved the crossed DNAs into a different channel containing 1 nM cohesin, 2.5 nM Scc2-Scc4, and 1 mM ATP (60 s, 50 mM NaCl) before returning the DNAs to a channel containing 1 mM ATP in 300 mM NaCl. We reversed the manipulation of DNA2, first moving bead 3 upward and over DNA1 before manipulating beads 3 and 4 so that DNA2 was rotated −90° and lowered back to the original position where DNA1 and DNA2 were parallel to each other. When we moved the beads to a channel containing SYTOX Orange to visualize DNA, we observed that DNA1 and DNA2 were bridged (fig. S8), as expected from our analysis of parallel DNA bridging in the dual-trap optical tweezers setup (Fig. 4B, Cohesin + Scc2/4 + 1 mM ATP). We then tested whether simultaneously moving DNA2 using beads 3 and 4 in the *x* axis would cause the sliding of the bridge along DNA1 (Fig. 4C). We observed that the bridge could be moved, showing that cohesin can slide on DNAs while tethering two DNA molecules in trans (Fig. 4C and movies S4 and S5). When we applied force to disrupt the bridge [moving bead 3 down in the *y* axis (away from beads 1 and 2); Fig. 4D], we were not able to break apart the cohesin tether. At high forces, the interaction between the ends of the DNAs and the beads often snapped (Fig. 4D and movie S6). Amazingly, cohesin bridges resisted this, and half of DNA2 could be observed hanging from the bridge (Fig. 4D and movie S6). We conclude that permanent intermolecular cohesin bridges can slide on DNA and resist high force. We predict that the forces exerted to disrupt the

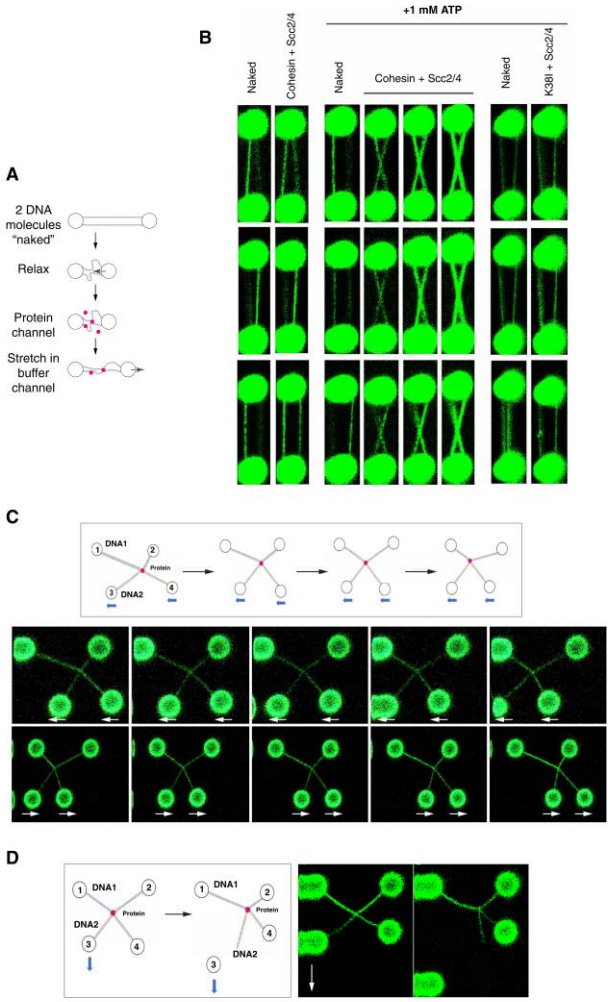



**Fig. 4. Cohesin and Scc2-Scc4 mediate intermolecular DNA bridges that slide on DNAs.** (**A**) Schematic representation of the experimental design for the dual-trap optical tweezer to generate permanent intermolecular cohesin bridges. Two λ-DNA molecules are tethered between the two beads and incubated in a relaxed position (3-μm bead distance) in the presence or absence of protein in buffer containing 50 mM NaCl. The relaxed molecules are then moved to a different channel containing 300 mM NaCl and reextended. Imaging is done before incubations and after reextension in a buffer containing 300 mM NaCl and 50 nM SYTOX Orange to visualize DNA. (**B**) Two λ-DNA molecules were tethered and treated as described in (A) and incubated with either (i) 1 nM cohesin, 2.5 nM Scc2-Scc4, and no ATP (Cohesin + Scc2/4, left); (ii) 1 nM cohesin, 2.5 nM Scc2-Scc4, and 1 mM ATP (Cohesin + Scc2/4, middle); or (iii) 1 nM cohesin ATPase mutant K38I, 2.5 nM Scc2-Scc4, and 1 mM ATP (K38I + Scc2/4, right). Imaging was performed before incubation and after DNA reextension in a buffer containing 300 mM NaCl to minimize DNA entanglement and 50 nM SYTOX Orange to visualize DNA. Images from three independent experiments are shown. Three independent experiments are shown for each category. (**C**) Schematic representation of the experimental design to test for sliding of permanent cohesin bridges (top diagram). Following the formation of an intermolecular cohesin bridge (see fig. S8 for details in bridge formation protocol), beads 3 and 4 were moved together in the *x* axis to slide the bridge along DNA1. Images showing two representative sliding experiments are shown. Experiments were performed in a buffer containing 300 mM NaCl and 50 nM SYTOX Orange. Movies of the experiments are shown in movies S4 and S5. The experiment was performed three times, and sliding was observed in all cases. (**D**) Schematic representation of the experimental design to disrupt intermolecular cohesin bridges. Following the formation of an intermolecular cohesin bridge, bead 3 is moved down in the *y* axis until one of the DNA ends loses contact with the bead. Imaging was performed before and after the pull in a buffer containing 300 mM NaCl and 50 nM SYTOX Orange. Representative experiment is shown. A movie of the experiment is shown in movie S6.

interaction between the DNAs and the bead exceed 80 pN. At these high forces, the prediction is that all the protein interfaces on cohesin rings should be disrupted. Therefore, cohesin association with DNA in permanent tethers is likely to occur in a manner that resists opening of the interfaces.

### Cohesin DNA-tethering activity is conserved in human cohesin tetramers containing Stag1

Previous studies using purified cohesin from different organisms did not report DNA bridging activities (*15*–*17*); however, the studies did not use budding yeast cohesin. We therefore decided to test whether the bridging activity observed is specific for *S. cerevisiae* cohesin tetramers or it has been conserved in cohesin from other organisms. To this aim, we purified the human cohesin (hCohesin) tetramer complex, containing hSmc1, hSmc3, hRad21, and Stag1, as described previously (fig. S9A) (*28*). We then tested whether hCohesin could bridge DNA intramolecularly. We captured a single λ-DNA molecule and generated an FE curve in the absence of protein to confirm the presence of naked DNA. We then moved the DNA to a channel containing 1 nM hCohesin and 1 mM ATP in 50 mM NaCl and incubated it for 30 s in a relaxed conformation (~3 μm between beads). We then moved the relaxed DNA to a channel without protein in the presence of 1 mM ATP in 125 mM NaCl. Reextending the DNA resulted in FE curves with a naked DNA profile (Fig. 5A, hCohesin), demonstrating that hCohesin on its own cannot promote DNA bridges. Although we could not obtain hScc2-Scc4, we decided to test whether the budding yeast loader complex Scc2-Scc4 (scScc2-Scc4) had any effect on hCohesin activity. To this aim, we repeated the intramolecular DNA bridging assays with hCohesin and included the Scc2-Scc4 loader complex in the incubations. Relaxed DNA was incubated in the presence of 1 nM hCohesin tetramer, 2.5 nM scScc2-Scc4 complex, and 1 mM ATP in 50 mM NaCl. The relaxed DNA was then moved to a channel with 1 mM ATP in 125 mM NaCl. Reextension yielded the sawtooth features characteristic of intramolecular bridge rupture events (Fig. 5B, hCohesin + Scc2/4) detected with yeast cohesin tetramers (Fig. 3C, 125 mM). Therefore, hCohesin tetramers containing Stag1 have conserved the ability to bridge DNA. hCohesin was able to form both reversible and permanent bridges (Fig. 4B, hCohesin + Scc2/4).

### Cohesin compacts linear DNA molecules stretched by very low forces

Besides mediating sister chromatid cohesion (*1*, *2*), cohesin holds individual chromatids in cis, thus forming loops (*4*, *29*, *30*). Recently, yeast condensin was the first SMC complex shown to exhibit an activity compatible with loop extrusion (*7*). It is unclear whether this activity is also present in the other eukaryotic SMC complexes cohesin and Smc5/6. Condensin can compact linear DNA against forces of up to 2 pN (*19*). However, condensin loop extrusion activity is only observed when DNA is stretched under significantly lower forces (below 0.5 pN) (*7*). We purified yeast condensin (fig. S9, B and C) using an established protocol (*20*, *31*) and tested whether, as predicted from studies using magnetic tweezers (*19*), it could also compact λ-DNA molecules extended in the optical tweezers against a force of 1 pN. A single λ-DNA molecule was first captured between the beads. We then immobilized one of the beads and applied a constant force of 1 pN to the other bead in the opposite direction. This maintains the DNA extended with ~14 μm between beads. We then moved the DNA to a channel containing 1 nM condensin in 50 mM NaCl buffer supplemented with 1 mM ATP. We incubated the extended DNA recording the distance between the two beads over time (Fig. 5C, Condensin). We observed progressive decrease of the distance between the beads (Fig. 5C, Condensin, and fig. S10), consistent with the activity of condensin as a motor that compacts DNA (*19*). Some condensation events occurred in short bursts and caused the molecule to shorten ~1 to 2 μm in a few seconds (Fig. 5C, Condensin, and fig. S10). After incubation, we generated an FE curve, which showed the presence of sawtooth peaks characteristic of protein-mediated DNA bridging (Fig. 5C, bottom) (*25*, *26*). Condensin bridges were fully reversible and disappeared when the DNA was extended (Fig. 5C, bottom). It is unclear whether the compaction observed is due to loop extrusion because this activity was reported to occur at forces below 1 pN (*7*). Next, we sought to text whether yeast cohesin tetramers could also compact extended λ-DNA molecules in this assay. We incubated the DNA extended using 1 pN of force with 1 nM cohesin, 2.5 nM Scc2-Scc4 complex, and 1 mM ATP in 50 mM NaCl buffer and incubated the extended DNA recording the distance between the two beads (Fig. 5D, Cohesin). The distance between the beads did not change over time (Fig. 5D, Cohesin, and fig. S10); therefore, we conclude that cohesin tetramers do not exhibit DNA compaction activity in this assay. As expected, the FE curve generated after incubation showed no evidence of protein-mediated DNA bridging (Fig. 5D, bottom), and similar results were obtained when we used a stretching force of 0.5 pN.

Since loop extrusion activity of condensin occurs at forces below 0.5 pN (*7*), we considered the possibility that yeast cohesin might also be able to extrude loops (and hence condense DNA) at extremely low forces. Below 0.5 pN, our optical tweezer did not reliably maintain the distance between the beads (data not shown). We therefore used single-tethered DNA curtains and different flow rates to extend DNA at very low tensions. Initially, we incubated cohesin in the presence of Scc2/Scc4 and ATP using a 125 mM NaCl buffer and a flow rate of 30 μl/min; however, we did not observe compaction of single-tethered DNAs (data not shown). We then decided to reduce the ionic strength of the buffer to 50 mM NaCl and the flow rate to 20 μl/min (Fig. 5E). We did not observe compaction in the course of the experiment (Fig. 5E). However, when we further reduced the flow rate to 10 μl/min we observed slow compaction of the majority of the molecules (Fig. 5F). Last, we performed the same experiment but stopped the flow after protein injection (Fig. 5G). We observed rapid compaction of the single-tethered DNAs (Fig. 5G). From these data, we conclude that DNA compaction in single-tethered DNA curtains at such low flow is likely to be formed as a consequence of compaction that might involve loop extrusion since this activity only occurs at low ionic strength conditions and when DNA is extended by very low force.

## DISCUSSION

The original role attributed to cohesin was the maintenance of sister chromatid cohesion from S phase until the anaphase onset (*1*, *2*). Here, we have developed powerful single-molecule assays to probe the mechanisms by which cohesin holds DNAs together. Using them, we have shown that cohesin complexes can form different types of bridges between dsDNAs and that this requires Scc2-Scc4 and ATP. The two classes of cohesin tethers exhibited different physical properties, particularly the sensitivity to being broken by force. The reversible bridges were disrupted when moderate forces (5 to 40 pN)

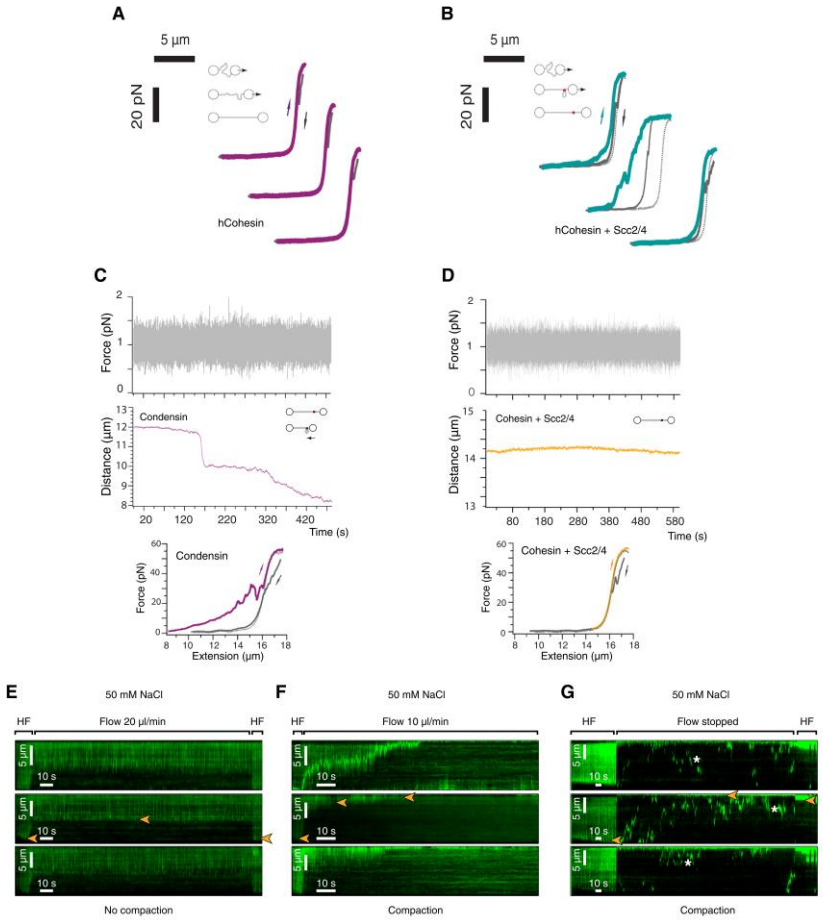



**Fig. 5. Cohesin does not compact linear DNA molecules stretched under low force.** (**A**) FE curve for λ-DNA preincubated with 1 nM human cohesin and 1 mM ATP in 125 mM NaCl buffer (hCohesin). Dotted line is fit to worm-like chain model. After capturing a single DNA molecule between two optically trapped beads, DNA was incubated in the presence of protein in 50 mM NaCl buffer in a relaxed conformation (3-µm bead distance) for 30 s and then moved to the 125 mM NaCl buffer channel for extension and measurements. No evidence of DNA bridges was observed under this condition. (**B**) FE curve for λ-DNA preincubated with 1 nM human cohesin, 2.5 nM yeast Scc2-Scc4, and 1 mM ATP in 125 mM NaCl buffer (hCohesin + Scc2/4). Experimental procedure as in (A). FE curves exhibited multiple rupture events indicating the presence of reversible and permanent DNA bridges. (**C**) DNA compaction trace for λ-DNA molecule extended using a force of 1 pN (top). The DNA was tethered between two beads. One bead was clamped (fixed), while a 1-pN force was applied to the second bead to maintain the molecule extended. The DNA was then incubated in the presence of 1 nM condensin (1 mM ATP in 50 mM NaCl) (left, magenta trace). The FE curve for the λ-DNA full extension after incubation is shown (bottom). Additional examples can be found in fig. S10. (**D**) DNA compaction trace for λ-DNA molecule extended using a force of 1 pN (top) in the presence of 1 nM cohesin and 2.5 nM Scc2-Scc4 complex (1 mM ATP in 50 mM NaCl) (right, yellow trace). The distance between the beads was recorded over time. The FE curve for the λ-DNA full extension after incubation is shown (bottom). Additional examples can be found in fig. S10. (**E**) Kymograms of single-tethered λ-DNA stained with (YOYO-1) during the incubation with yeast cohesin and Scc2-Scc4 in the presence of ATP in 50 mM NaCl buffer at a flow rate of 20 µl/min. HF, high flow. The free end of DNA is marked with orange arrowheads. No compaction of single-tethered λ-DNAs was observed. (**F**) Kymograms of single-tethered λ-DNA stained with (YOYO-1) during the incubation with yeast cohesin and Scc2-Scc4 in the presence of ATP in 50 mM NaCl buffer at a flow rate of 10 µl/min. The conditions are as in (E) except for the reduced flow rate. Slow compaction of single-tethered λ-DNAs was observed over time (orange arrowheads mark the free end of DNA). (**G**) Kymograms of single-tethered λ-DNA stained with (YOYO-1) during the incubation with yeast cohesin and Scc2-Scc4 in the presence of ATP in 50 mM NaCl buffer at stopped flow. The free end of DNA is marked with orange arrowheads. The HF phase at the end of the experiment shows that the DNA was compacted during the stopped flow phase. Note that under stopped flow conditions, DNA molecules that diffuse laterally on the flow chip can transiently cross the field of view and also appear in a kymogram representation. Examples are marked with asterisks (*). These events bear no relevance for the interpretations of the assay.

were applied (Fig. 3C). In contrast, permanent bridges could withstand extreme forces without being disrupted (Figs. 3C and 4D). They are also more predominant in high ionic strength conditions (Fig. 3C). On the basis of these physical properties, we propose that permanent bridges represent cohesin complexes that maintain sister chromatid cohesion. However, further characterization of their genesis, architecture, and biochemistry will be important to confirm such proposal. Reversible bridges were more predominant at low-salt concentrations (Fig. 3C), which suggest that they are likely formed by protein-protein interactions. In low salt, cohesin is likely to be saturated on DNA and being relatively sticky could easily engage in nonspecific interactions. Therefore, some reversible bridging events could potentially

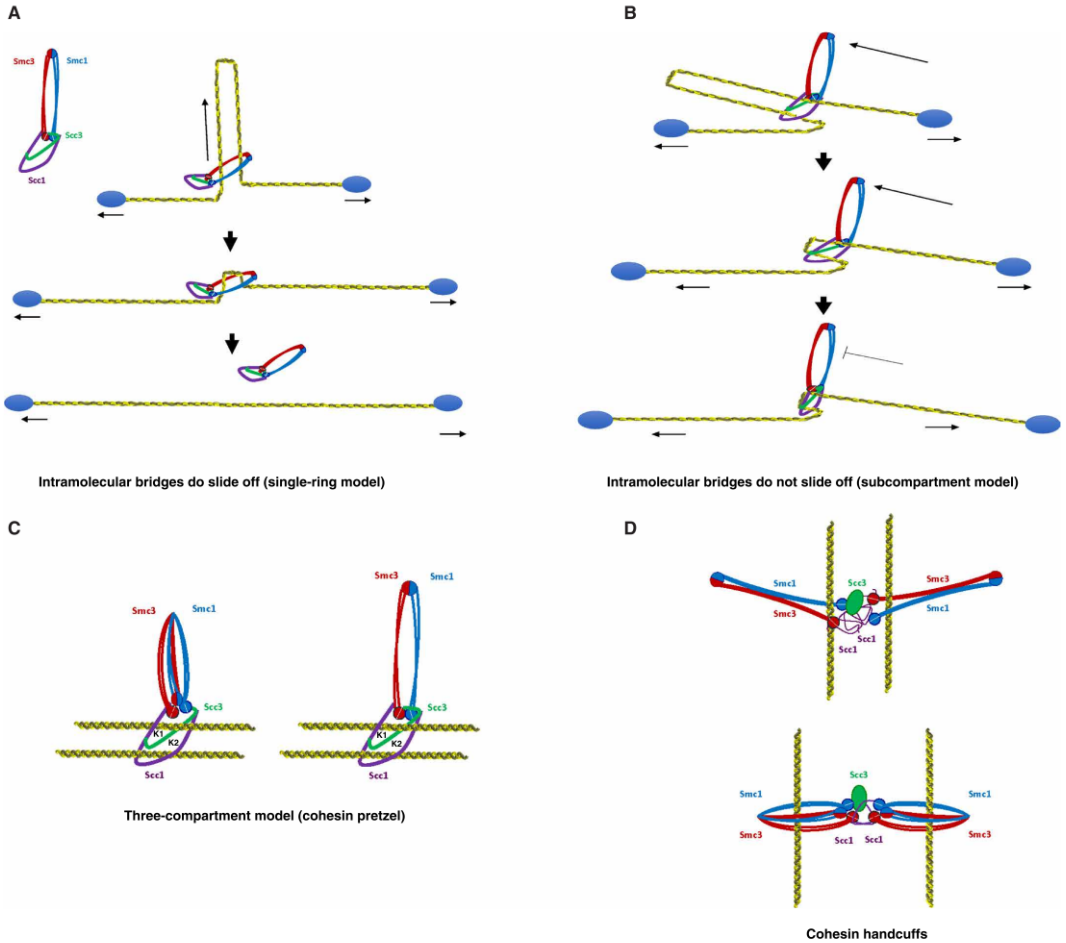



**Fig. 6. Tentative models for permanent cohesin bridges.** (**A**) Schematic representation of expected behavior of intramolecular cohesin tethers from the previously proposed ring model. The model proposes that cohesin co-entraps two DNAs within its ring structure, i.e., both DNAs occupy one physical space within cohesin. From this model, it is expected that cohesin should be fully displaced from λ-DNA molecules when tethering in cis as force is applied to separate the beads. This is not what it was observed experimentally (Fig. 3C and fig. S5). (**B**) Schematic representation of expected behavior of intramolecular cohesin tethers from the subcompartment model. The subcompartment model is based on the assumption that DNAs are located in different physical compartments. The prediction from the model is that cohesin cannot be fully displaced from λ-DNA molecules when tethering them in cis. This is what we observed experimentally (Fig. 3C and fig. S5). (**C**) Proposed model for a single cohesin complex with at least three subcompartments (cohesin pretzel). In this model, sister DNAs occupy two different chambers (K1 and K2) of the K (kleisin) compartment formed between the SMC ATPase heads and the Scc1 subunit (*36*). Two possible conformations of SMC hinges are shown. Note that the experimental data are also compatible with the possibility that both DNAs jointly travel through the two chambers (K1 and K2) of the K compartment. (**D**) Schematic representation of previously proposed cohesin handcuffs models holding sister DNAs in different compartments of two separate complexes, which also fits with our experimental observations.

represent nonspecific protein aggregation. In particular, this might be the case for intramolecular DNA bridging at 50 mM NaCl salt (Fig. 3C). However, even under these conditions, reversible bridges were ATP and Scc2/4 dependent (fig. S3A). At 125 mM NaCl salt, which is in the physiological range, reversible bridges were also significant and resisted forces of up to 40 pN (Fig. 3C), strongly arguing that reversible bridges are biologically relevant. Previous studies have demonstrated that cohesin can use nontopological mechanisms (*32*); in addition, interallelic complementation between different cohesin alleles has also been reported (*33*). It is therefore possible that reversible DNA bridges reflect functional cohesin-cohesin interactions.

Recent studies have interrogated cohesin mechanisms using biochemical reconstitution of topological loading onto plasmids (*14*, *18*, *34*, *35*). We believe that the single-molecule assay presented in this study is more informative for the study of cohesin bridging. In our hands, cohesin loading in the gel-based assay was not strictly ATP dependent and was not stimulated by Scc2/4, as observed for *S. pombe* cohesin (*14*, *18*, *34*, *35*). Topological loading efficiency can be dependent on multiple factors, but critically on the amount of protein used, the times of incubation, and the number and stringency of the washes. We followed the original protocol described for *S. pombe* cohesin (*18*), and despite attempting different conditions, we never observed ATP-dependent loading. It is therefore likely that *S. pombe* and *S. cerevisiae* cohesins behave differently. The observation that *S. pombe* cohesin does not show bridging activity in double-tethered DNA curtains (*15*), while *S. cerevisiae* cohesin does (Fig. 2, E and F), supports this possibility.

Our results using two DNA molecules demonstrate that permanent cohesin tethers can slide when force is applied (Fig. 4C); however, when the permanent bridges occur in cis, cohesin complexes cannot slide off the DNA molecules (Fig. 3C and fig. S5). The simplest explanation is that the two DNA molecules tethered are not located in the same physical space within the protein (Fig. 6A). The two main models proposed to explain how cohesin holds sister chromatids are the "ring" and "handcuff" models. The basic difference between these two models is the fact that in the ring model, the two DNAs occupy the same physical space within cohesin, i.e., they are co-entrapped in one compartment of the cohesin structure (*10*, *11*), while in the handcuff model (and all its variations), the two DNAs are located in different physical compartments (*1*, *12*, *13*), generally argued to be two separate (but interacting) complexes. On the basis of the single-ring model, it would be expected that cohesin slides off DNA molecules when bridging them intramolecularly (Fig. 6A). In contrast, our observations suggest that this is not the case (Fig. 3C). Using in vivo cysteine crosslinking of trimer cohesin complexes, it has been recently shown that cohesin has different subcompartments (*36*). Sister DNAs occupied the K (kleisin) compartment formed between the SMC ATPase heads and the Scc1 subunit (*36*). Scc3 and Scc1 form a module that binds DNA and is necessary for cohesin association to chromosomes (*37*), but Scc3 was not crosslinked in the subcompartment study (*36*). We propose that DNAs in permanent cohesin bridges might be held in two chambers of the K (kleisin) compartment (Fig. 6B, K1 and K2), physically separated by Scc3 (Fig. 6B), and the architecture would resemble a "pretzel-like" structure (Fig. 6B). The DNAs might be separated (one in K1 and the other in K2), or might travel through the two K compartments together. Alternatively, different compartments of two cohesin complexes might be involved (Fig. 6C).

Kimura *et al.* (*5*) first proposed that the SMC complex condensin might generate DNA loops (*5*). This was conceived as one of two models that could explain how condensin specifically produced (+) trefoil knots in the presence of a type II topoisomerase (*5*). The proposal was based on an earlier model of "loop expansion" that was put forward for bacterial MutS action (*38*). MutS loop expansion was shown to occur as a consequence of ATP-dependent bidirectional movement of the MutS dimer from the initial loading site (*38*). The proposal of Kimura *et al.* (*5*) has been recently demonstrated directly through the observation of condensin-dependent DNA loop–like structures on surface-tethered, flow-stretched DNA (*7*). The loop extrusion activity of cohesin was also conceived as a model that could explain the role of cohesin in genome folding through cis looping and the preferential use of convergent CTCF DNA motifs at TAD borders (*8*, *9*). We detected DNA compaction by yeast cohesin tetramers at very low flow rates (Fig. 5, F and G), as would be predicted from a loop extrusion activity similar to the one shown for condensin (*7*). HiC data show that removal of cohesin leads to loss of contacts at TAD boundaries (*6*, *8*, *39*), demonstrating that the complex is involved in the formation or maintenance of loops. It is likely that cohesin extrudes DNA loops in a similar manner to condensin (*7*). However, our data, although consistent with cohesin function as a loop extruder, do not demonstrate it. We would like to note that our data showing intramolecular tethering by cohesin do not imply that cohesin generates loops in vivo through random DNA bridging. We feel that this would be highly unlikely. The intramolecular tethers observed might reflect an in vitro activity (as cohesin is unlikely to differentiate between cis and trans tethering when loaded onto DNA in these assays). Further experiments will be required to test whether intramolecular tethering is of any relevance in vivo. The activities described here are fully consistent with the original role attributed to cohesin in maintaining sister chromatid cohesion (*1*, *2*). Our work provides a new critical tool for future investigations to further decipher how cohesin executes one of the critical functions required for genome inheritance, i.e., maintaining sister chromatids in close proximity from the time they are born in S phase until they are separated in anaphase.

## MATERIALS AND METHODS

### Protein expression and purification

The different subunits of the *S. cerevisiae* Scc2-Scc4 and cohesin complexes were synthesized under the control of galactose-inducible promoters and cloned into multicopy episomal vectors (*URA3-SCC4-GAL1-10promoter-SCC2*-3xmyc-3xStrepII;*TRP1-SMC1*-3xStrepII-*GAL1-10promoter-SMC3 GAL7promoter-MCD1*-8xHis-3xHA; *URA3-GAL1-10promoter-SCC3*). Yeast W303-1a strains carrying the different constructs (CCG14800 for the Scc2-Scc4 complex, CCG14801 for cohesin tetramer, and CCG14815 for cohesin *smc3*-K38I tetramer) were grown at 30°C in selective dropout medium containing 2% raffinose and 0.1% glucose to an $OD_{600}$ (optical density at 600 nm) of 1. Protein expression was induced by addition of 2% galactose, and cells were grown for further 16 hours at 20°C. Cells were then harvested by centrifugation at 4°C, resuspended in two volumes of buffer A [25 mM Hepes (pH 7.5), 200 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol] containing 1× cOmplete EDTA-free protease inhibitor mix (Roche), frozen in liquid nitrogen, and lysed in Freezer-Mill (SPEX CertiPrep 6870). Cell powder was thawed at 4°C for 2 hours before mixing it with one volume of buffer A containing benzonase (Millipore) and incubated at 4°C for an extra hour. Cell lysates were clarified by centrifugation at 45,000*g* for 1 hour followed by filtration using 0.22-μm syringe filters.

Clarified lysates were loaded onto 5-ml StrepTrap-HP columns (GE Healthcare) preequilibrated with buffer A. The resin was washed with five column volumes of buffer A and eluted with buffer B (buffer A containing 5 mM desthiobiotin). The peak fractions containing the overexpressed proteins were pooled together, and salt concentration was adjusted to 150 mM NaCl using 100 mM NaCl buffer A. Samples were then filtered as described above to remove residual aggregates and loaded onto 5-ml HiTrap Heparin HP (GE Healthcare) columns preequilibrated with 150 mM NaCl buffer A. Elution was carried out using a linear gradient from 150 mM to 1 M NaCl in buffer A. Peak fractions were pooled and concentrated by centrifugal ultrafiltration (100 kDa Amicon Ultra, Millipore). Salt concentration was adjusted to 300 mM NaCl during the concentration step. Gel filtration was carried out using a Superose 6 Increase 100/300 GL column (GE Healthcare) in 300 mM NaCl buffer A. Fractions corresponding to monomeric complexes were pooled and concentrated as described above. Purified proteins were analyzed by SDS-PAGE (NuPAGE 4 to 12% bis-tris protein gels, Thermo Fisher Scientific) and Coomassie staining (InstantBlue, Expedeon). Protein identification was carried out by mass spectrometry analysis and Western blot. *S. cerevisiae* condensin complex was expressed and purified, as previously described (*20*, *31*).

Human cohesin tetramer was purified, as described before (*28*). Human cohesin subunits (RAD21, SMC1A, SMC3-FLAG, 10xHis-SA1) were coexpressed in High Five insect (BTI-Tn-5B1-4) cells. Cells were disrupted by short sonication. Afterwards, the lysate was clarified by high-speed centrifugation. The complex was then purified via HisTrap [washing buffer: 25 mM tris (pH 7.5), 500 mM NaCl, 5% glycerol, 2 mM $MgCl_2$, 20 mM imidazole, 0.01% Tween-20, 20 mM β-mercaptoethanol; elution buffer: 25 mM tris (pH 7.5), 150 mM NaCl, 5% glycerol, 2 mM $MgCl_2$, 150 mM imidazole, 0.01% Tween-20]. Fractions were pooled and dialyzed [25 mM tris (pH 7.5), 150 mM NaCl, 5% glycerol, 2 mM $MgCl_2$]. The protein was further purified by tandem ion exchange chromatography by using an anion-exchange column connected to a cation exchange column. The complex was then eluted from the cation-exchange column [25 mM tris (pH 7.5), 1 M NaCl, 5% glycerol, 2 mM $MgCl_2$]. Subsequently, the peak fractions were pooled and dialyzed into storage buffer [25 mM tris (pH 7.5), 150 mM NaCl, 5% glycerol, 2 mM $MgCl_2$]. Purity was confirmed by gel electrophoresis and mass spectrometry.

### Electrophoretic gel mobility shift assay

Increasing concentrations of the Scc2-Scc4 complex ranging from 100 to 800 nM were incubated for 45 min with 50 ng of pUC19 at 30°C in 25 mM tris-HCl (pH 7.0), 50 mM NaCl, 8% glycerol, bovine serum albumin (BSA; 0.1 mg/ml), and 0.5 mM dithiothreitol (DTT) in a final volume of 15 μl. The reactions were resolved by electrophoresis for 1 hour at 80 V on 0.8% (w/v) tris-acetate-EDTA (TAE) agarose gels at 4°C. DNA was detected on a fluorescent image analyzer (FLA-5000, Fujifilm) after SYBR Green I (Invitrogen, Thermo Fisher Scientific) gel staining. Condensin assays were carried out as previously described (*20*).

### Protein cross-linking and electron microscopy

For cross-linking of cohesin complex, protein samples were incubated with BS3 at a 1:3000 molar ratio in buffer XL [25 mM Hepes, 125 mM NaCl, 5% glycerol, 1 mM DTT (pH 8)] for 2 hours on ice before quenching with 10 mM tris-HCl (pH 8) for 30 min on ice.

Negative-stain grids were prepared as follows: 3.5 μl of suspended sample (final concentration of 0.02 mg/ml in buffer XL) was deposited on glow-discharged grids coated with a continuous carbon film. The sample was left on the grid for 1 min before blotting the excess liquid. A 3.5-μl drop of 2% uranyl acetate solution was added for 1 min, the stain was blotted away, and the grids were left to dry.

A set of 250 micrographs was collected on a Philips CM200 TWIN FEG electron microscope operated at 160 kV. Images were recorded on a Tietz 2k charge-coupled device camera at a nominal magnification of 38,000 and a final pixel size of 3.58 Å, contrast transfer function (CTF) parameters were estimated using Gctf (*40*). A total of ~9000 particles were automatically picked using Gautomatch software using class averages obtained from a manually picked subset of 1500 particles as references. The following two-dimensional classifications were performed with RELION v3.0 beta (*41*).

### In vitro cohesin loading assay

Cohesin loading assays were done as described in (*14*) using the pUC19 plasmid. Topologically bound DNA-cohesin complexes were immunoprecipitated using a μMACS HA Isolation kit (Miltenyi Biotec). Following incubation with Pst I and/or protein digestion, the recovered DNA was analyzed by electrophoresis on a 0.8% (w/v) TAE agarose gel in 1× TAE and visualized as described above.

### ATPase assays

For the ATPase assays, 30 nM cohesin was incubated at 29°C with 60 nM Scc2/4 and 0.2 nM λ-DNA (New England Biolabs) in ATPase buffer [35 mM tris-HCl (pH 7.0), 20 mM NaCl, 0.5 mM $MgCl_2$, 13.3% glycerol, 0.003% Tween-20, 1 mM tris(2-carboxyethyl) phosphine (TCEP), BSA (0.2 mg/ml)]. The reaction was started by adding 400 μM ATP spiked with [γ-$^{32}$P]ATP. One microliter of samples was taken after 1, 15, 30, and 60 min. The reaction was immediately stopped by adding 1 μl of 50 mM EDTA before spotting the samples on polyethyleneimine cellulose F sheets. The free phosphate was separated from ATP using thin-layer chromatography with 0.5 M LiCl, 1 M formic acid as the mobile phase. The spots were detected on a phosphor imager and analyzed using ImageJ. Data points were corrected for spontaneous ATP hydrolysis. Each reaction was performed in triplicate. Data were fitted to Michaelis-Menten kinetics.

### Single-molecule experiments

DNA curtain experiments were performed as described previously (*42*). Briefly, flow cells were produced by deposition of chromium features onto fused silica microscope slides by e-beam lithography. Flow cells were connected to a microfluidics system based on a syringe pump (Landgraf GmbH) and two injection valves (Idex) and illuminated by 488- or 561-nm lasers (Coherent) in a prism-type total internal reflection fluorescence (TIRF) configuration on an inverted microscope (Nikon Ti2e). Imaging was performed using an electron multiplying charge-coupled device (EMCCD) camera (Andor iXon life) with illumination times of 100 ms. λ-DNA (NEB) was end-modified by hybridizing biotinylated or digoxigeninated oligos complementary to the cos site and purified by size exclusion chromatography. Modified λ-DNA was anchored to the surface of a lipid bilayer in flow cells by biotin-streptavidin-biotin interactions, stretched by flow across chromium barriers, and anchored to downstream chromium pedestals by the digoxigenin-binding protein DIG10.3 (*43*). Experiments were performed in buffer M [40 mM tris-HCl (pH 7.8), 1 mM $MgCl_2$, 1 mM DTT, BSA (1 mg/ml), 0.16 nM YOYO-1].

Cohesin complexes were labeled by incubating them at a concentration of 3 nM in a small volume of buffer M supplemented with 50 mM NaCl with 3× molar excess Qdots (SiteClick 705 kit, Invitrogen) fused to anti-HA antibodies (3F10, Roche) for 30 min at 4°C. The mixture was then supplemented with 8 nM Scc2/Scc4, 100 μm biotin, and 0.5 mM nucleotide (ATP, ADP, or ATPγS), if required, before injection. For diffusion measurements, the flow cell was flushed after the completion of loading with buffer M supplemented with KCl at the indicated concentrations and the flow was stopped. Illuminations were performed either continuously (diffusion and lifetime measurements) or with lower frame rates (intermolecular bridging videos). To minimize photodamage, 488-nm pulses to illuminate the DNA, if required, were only used at every 10th illumination.

Videos were recorded in NIS Elements (Nikon) and analyzed using custom-written software in Igor Pro (WaveMetrics). Lifetime measurements and initial binding distributions of cohesin complexes on DNA were generated by manually analyzing kymograms. Survival curves were generated by a Kaplan-Meier estimator, bootstrapped, and fitted to a double-exponential model.

For the determination of diffusion coefficients, labeled cohesin complexes were tracked using custom-written software, and the diffusion coefficients were extracted using a maximum-likelihood estimator (*44*), as described previously (*15*).

Optical tweezers experiments were carried out on C-trap and Q-trap systems integrating optical tweezers, confocal fluorescence microscopy, and microfluidics and recorded using BlueLake software (LUMICKS). The laminar flow cell was passivated using 0.50% pluronic and BSA (2 mg/ml). Biotin-labeled double-stranded λ-DNA molecules were tethered between two streptavidin-coated polyesterene beads (4.42 μm in diameter, Spherotech). Depending on the experiment, one or two individual double-stranded λ-DNA molecules were attached between two beads. The beads were previously passivated with BSA (1 mg/ml). After DNA capture, beads were incubated inside the protein channel either in a relaxed (~3 μm apart) or extended position (force clamp at 5 pN, ~14 μm apart) for 30 s and then returned to the buffer channel for FE, force clamp, and fluorescence analysis. Cohesin and Scc2-Scc4 complex were used at 1 nM and 2.5 nM concentrations, respectively. Beads and DNA catching and protein loading were performed in a buffer containing 50 mM tris-HCl (pH 7.5), 50 mM NaCl, 2.5 mM $MgCl_2$, BSA (0.5 mg/ml), 40 μM biotin, and 1 mM DTT. When indicated, ADP, ATPγS, and ATP were added to both protein and buffer channels at a final concentration of 1 mM. Salt concentration was modified from 50 mM to 125, 300, or 500 mM NaCl in the buffer channel as specified in the text and figures. FE curves were performed at a speed of 1 μm/s. Compaction experiments were carried out at a constant force of 1 pN. For friction experiments, beads were moved 6 μm, back and forth, at a speed of 0.2 μm/s. SYTOX Orange (Invitrogen, Thermo Fisher Scientific) was used at a final concentration of 50 mM for DNA imaging, using a 532-nm wavelength laser. Force data were processed using Igor Pro 7 software (WaveMetrics), and images were processed using Adobe Photoshop CC.

### Western blot

For Western blot, 2 μg of purified complexes was run on NuPAGE 4 to 12% bis-tris gels (Thermo Fisher Scientific), transferred to Immobilon-P membranes (Millipore), and probed with anti-Strep (ab180957, Abcam, 1:5000) and anti-HA (3F10, Roche, 1:5000) antibodies in 5% milk–phosphate-buffered saline (PBS)/0.01% Tween overnight at 4°C. Membranes were then washed and incubated with horseradish peroxidase anti-rabbit (Santa Cruz Biotechnology, 1:40,000) and anti-rat (Jackson ImmunoResearch, 1:10,000) secondary antibodies, respectively, for 1 hour at room temperature. Immunoblots were developed using the Luminata Forte detection reagent (Millipore) and Hyperfilms ECL (GE Healthcare).

### Liquid chromatography–tandem mass spectrometry

Samples were processed by in-Stage Tip digestion (PreOmics GmbH, Planegg/Martinsried) following the manufacturer's recommendation. Protein digests were solubilized in 30 μl of reconstitution buffer and transferred to autosampler vials for liquid chromatography–mass spectrometry analysis. Peptides were separated using an Ultimate 3000 RSLC nano–liquid chromatography system (Thermo Fisher Scientific) coupled to an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) via an EASY-Spray source. Sample volumes were loaded onto a trap column (Acclaim PepMap 100 C18, 100 μm × 2 cm) at 8 μl/min in 2% acetonitrile and 0.1% trifluoroacetic acid. Peptides were eluted online to an analytical column (EASY-Spray PepMap C18, 75 μm × 50 cm). Peptides were separated using a ramped 120-min gradient from 1 to 42% buffer B [buffer A: 5% dimethyl sulfoxide (DMSO), 0.1% formic acid; buffer B: 75% acetonitrile, 0.1% formic acid, 5% DMSO]. Eluted peptides were analyzed operating in positive polarity using a data-dependent acquisition mode. Ions for fragmentation were determined from an initial MS1 survey scan at 30,000 resolution [at mass/charge ratio ($m/z$) of 200] in the Orbitrap followed by CID (collision-induced dissociation) of the top 10 most abundant ions in the Ion Trap. MS1 and MS2 scan AGC targets were set to $1 \times 10^6$ and $1 \times 10^5$ for a maximum injection time of 50 and 110 ms, respectively. A survey scan $m/z$ range of 350 to 1500 $m/z$ was used, with CID parameters of isolation width 1.0 $m/z$, normalized collision energy of 35%, activation Q of 0.25, and activation time of 10 ms.

Data were processed using the MaxQuant software platform (v1.6.2.3) with database searches carried out by the in-built Andromeda search engine against the UniProt *S. cerevisiae* database (6729 entries, v.20180305). A reverse decoy database was created, and results were displayed at a 1% false discovery rate for peptide spectrum matches and protein identification. Search parameters included the following: trypsin, two missed cleavages, fixed modification of cysteine carbamidomethylation and variable modifications of methionine oxidation, asparagine deamidation, and protein N-terminal acetylation. Label-free quantification (LFQ) was enabled with an LFQ minimum ratio count of 2. "Match between runs" function was used with match and alignment time limits of 0.7 and 20 min, respectively. Protein and peptide identification and relative quantification outputs from MaxQuant were further processed in Microsoft Excel, with hits to the "reverse database," "potential contaminants" (peptide list only), and "only identified by site" fields removed.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/11/eaay6804/DC1

Fig. S1. Topological loading of yeast cohesin on plasmid DNA.
Fig. S2. Analysis of yeast cohesin on DNA curtains.
Fig. S3. Intramolecular cohesin bridging requires ATP.
Fig. S4. Purification of budding yeast cohesin ATPase mutant.
Fig. S5. Permanent cohesin bridges are not displaced by physical stretching of λ-DNA.
Fig. S6. Cohesin does not capture two λ-DNAs in sequential steps.
Fig. S7. DNA friction experiments confirm the presence of cohesin complexes on extended λ-DNA.

Fig. S8. Generation of permanent cohesin bridges using a quadrupole-trap optical tweezer.
Fig. S9. Purification of human cohesin and yeast condensin.
Fig. S10. Budding yeast condensin, but not cohesin, compacts λ-DNA against 1 pN stretching force.
Table S1. Mass spectrometry analysis of cohesin wild type and ATPase mutant (Smc3-K38I) tetramer complexes and the loader complex Scc2-Scc4.
Table S2. Mass spectrometry analysis of cohesin ATPase mutant (Smc3-K38I) tetramer peptides showing peptides containing the K38I mutation for SMC3.
Movies S1 to S3. Time-lapse videos showing cohesin tethering.
Movies S4 and S5. Time-lapse videos showing sliding of intermolecular bridges in a quadruple-trap optical tweezer.
Movie S6. Time-lapse video showing pulling on intermolecular bridges in a quadruple-trap optical tweezer.

**View/request a protocol for this paper from *Bio-protocol*.**

## REFERENCES AND NOTES


1. V. Guacci, D. Koshland, A. Strunnikov, A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. *Cell* **91**, 47–57 (1997).
2. C. Michaelis, R. Ciosk, K. Nasmyth, Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, 35–45 (1997).
3. K. Shintomi, T. Hirano, The relative ratio of condensin I to II determines chromosome shapes. *Genes Dev.* **25**, 1464–1469 (2011).
4. S. S. P. Rao, S. C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K. R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, E. L. Aiden, Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
5. K. Kimura, V. V. Rybenkov, N. J. Crisona, T. Hirano, N. R. Cozzarelli, 13S condensin actively reconfigures DNA by introducing global positive writhe: Implications for chromosome condensation. *Cell* **98**, 239–248 (1999).
6. E. Alipour, J. F. Marko, Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* **40**, 11202–11212 (2012).
7. M. Ganji, I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, C. Dekker, Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).
8. G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, L. A. Mirny, Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
9. S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
10. S. Gruber, C. H. Haering, K. Nasmyth, Chromosomal cohesin forms a ring. *Cell* **112**, 765–777 (2003).
11. K. Nasmyth, How might cohesin hold sister chromatids together? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 483–496 (2005).
12. N. Zhang, S. G. Kuznetsov, S. K. Sharan, K. Li, P. H. Rao, D. Pati, A handcuff model for the cohesin complex. *J. Cell Biol.* **183**, 1019–1031 (2008).
13. I. Onn, J. M. Heidinger-Pauli, V. Guacci, E. Ünal, D. E. Koshland, Sister chromatid cohesion: A simple concept with a complex reality. *Annu. Rev. Cell Dev. Biol.* **24**, 105–129 (2008).
14. Y. Murayama, C. P. Samora, Y. Kurokawa, H. Iwasaki, F. Uhlmann, Establishment of DNA-DNA interactions by the cohesin ring. *Cell* **172**, 465–477.e15 (2018).
15. J. Stigler, G. Ö. Çamdere, D. E. Koshland, E. C. Greene, Single-molecule imaging reveals a collapsed conformational state for DNA-bound cohesin. *Cell Rep.* **15**, 988–998 (2016).
16. M. Kanke, E. Tahara, P. J. Huis in't Veld, T. Nishiyama, Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *EMBO J.* **35**, 2686–2698 (2016).
17. I. F. Davidson, D. Goetz, M. P. Zaczek, M. I. Molodtsov, P. J. Huis in 't Veld, F. Weissmann, G. Litos, D. A. Cisneros, M. Ocampo-Hafalla, R. Ladurner, F. Uhlmann, A. Vaziri, J.-M. Peters, Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.* **35**, 2671–2685 (2016).
18. Y. Murayama, F. Uhlmann, Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**, 367–371 (2014).
19. J. M. Eeftens, S. Bisht, J. Kerssemakers, M. Kschonsak, C. H. Haering, C. Dekker, Real-time detection of condensin-driven DNA compaction reveals a multistep binding mechanism. *EMBO J.* **36**, 3448–3457 (2017).
20. T. Terakawa, S. Bisht, J. M. Eeftens, C. Dekker, C. H. Haering, E. C. Greene, The condensin complex is a mechanochemical motor that translocates along DNA. *Science* **358**, 672–676 (2017).
21. N. J. Petela, T. G. Gligoris, J. Metson, B.-G. Lee, M. Voulgaris, B. Hu, S. Kikuchi, C. Chapard, W. Chen, E. Rajendra, M. Srinivisan, H. Yu, J. Löwe, K. A. Nasmyth, Scc2 is a potent activator of Cohesin's ATPase that promotes loading by binding Scc1 without Pds5. *Mol. Cell* **70**, 1134–1148 e7 (2018).
22. M. Minamino, T. L. Higashi, C. Bouchoux, F. Uhlmann, Topological in vitro loading of the budding yeast cohesin ring onto DNA. *Life Sci. Alliance* **1**, e201800143 (2018).
23. F. Bürmann, B.-G. Lee, T. Than, L. Sinn, F. J. O'Reilly, S. Yatskevich, J. Rappsilber, B. Hu, K. Nasmyth, J. Löwe, A folded conformation of MukBEF and cohesin. *Nat. Struct. Mol. Biol.* **26**, 227–236 (2019).
24. I. Brouwer, G. Sitters, A. Candelli, S. J. Heerema, I. Heller, A. J. de Melo, H. Zhang, D. Normanno, M. Modesti, E. J. Peterman, G. J. Wuite, Sliding sleeves of XRCC4–XLF bridge DNA and connect fragments of broken DNA. *Nature* **535**, 566–569 (2016).
25. N. Laurens, R. P. C. Driessen, I. Heller, D. Vorselen, M. C. Noom, F. J. H. Hol, M. F. White, R. T. Dame, G. J. L. Wuite, Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat. Commun.* **3**, 1328 (2012).
26. J. L. Syrjänen, I. Heller, A. Candelli, O. R. Davies, E. J. Peterman, G. J. Wuite, L. Pellegrini, Single-molecule observation of DNA compaction by meiotic protein SYCP3. *eLife* **6**, e22582 (2017).
27. I. Brouwer, G. A. King, I. Heller, A. S. Biebricher, E. J. G. Peterman, G. J. L. Wuite, Probing DNA–DNA interactions with a combination of quadruple-trap optical tweezers and microfluidics. *Methods Mol. Biol.* **1486**, 275–293 (2017).
28. F. Weissmann, G. Petzold, R. VanderLinden, P. J. Huis in 't Veld, N. G. Brown, F. Lampert, S. Westermann, H. Stark, B. A. Schulman, J.-M. Peters, biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2564–E2569 (2016).
29. L. Lazar-Stefanita, V. F. Scolari, G. Mercy, H. Muller, T. M. Guérin, A. Thierry, J. Mozziconacci, R. Koszul, Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J.* **36**, 2684–2697 (2017).
30. S. A. Schalbetter, A. Goloborodko, G. Fudenberg, J.-M. Belton, C. Miles, M. Yu, J. Dekker, L. Mirny, J. Baxter, SMC complexes differentially compact mitotic chromosomes according to genomic context. *Nat. Cell Biol.* **19**, 1071–1080 (2017).
31. J. St-Pierre, M. Douziech, F. Bazile, M. Pascariu, É. Bonneil, V. Sauvé, H. Ratsima, D. D'Amours, Polo kinase regulates mitotic chromosome condensation by hyperactivation of condensin DNA supercoiling activity. *Mol. Cell* **34**, 416–426 (2009).
32. M. Srinivasan, J. C. Scheinost, N. J. Petela, T. G. Gligoris, M. Wissler, S. Ogushi, J. E. Collier, M. Voulgaris, A. Kurze, K.-L. Chan, B. Hu, V. Costanzo, K. A. Nasmyth, The cohesin ring uses its hinge to organize dna using non-topological as well as topological mechanisms. *Cell* **173**, 1508–1519.e18 (2018).
33. T. Eng, V. Guacci, D. Koshland, Interallelic complementation provides functional evidence for cohesin-cohesin interactions on DNA. *Mol. Biol. Cell* **26**, 4224–4235 (2015).
34. G. Çamdere, V. Guacci, J. Stricklin, D. Koshland, The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. *eLife* **4**, e11315 (2015).
35. G. Ö. Çamdere, K. K. Carlborg, D. Koshland, Intermediate step of cohesin's ATPase cycle allows cohesin to entrap DNA. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9732–9737 (2018).
36. C. Chapard, R. Jones, T. van Oepen, J. C. Scheinost, K. Nasmyth, Sister DNA entrapment between juxtaposed smc heads and kleisin of the cohesin complex. *Mol. Cell* **75**, 224–237.e5 (2019).
37. Y. Li, K. W. Muir, M. W. Bowler, J. Metz, C. H. Haering, D. Panne, Structural basis for Scc3-dependent cohesin recruitment to chromatin. *eLife* **7**, e38356 (2018).
38. D. J. Allen, A. Makhov, M. Grilley, J. Taylor, R. Thresher, P. Modrich, J. D. Griffith, MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* **16**, 4467–4476 (1997).
39. A. L. Sanborn, S. S. P. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, E. L. Aiden, Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6456–E6465 (2015).
40. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
41. S. H. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
42. E. C. Greene, S. Wind, T. Fazio, J. Gorman, M.-L. Visnapuu, DNA curtains for high-throughput single-molecule optical imaging. *Methods Enzymol.* **472**, 293–315 (2010).
43. C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, D. Baker, Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013).
44. C. L. Vestergaard, P. C. Blainey, H. Flyvbjerg, Optimal estimation of diffusion coefficients from single-particle trajectories. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **89**, 022726 (2014).


**Acknowledgments:** We thank J. C. Danes and J. Andrecka (LUMICKS) for technical help. We thank our laboratory members for discussion and critical reading of the manuscript. We thank D. D'Amours, C. Haering, and J. Peters for sharing plasmids for the expression of yeast condensin and human cohesin. **Funding:** The work in the L.A. laboratory was supported by Wellcome Trust Senior Investigator award to L.A. (100955, "Functional dissection of mitotic

chromatin") and the "London Institute of Medical Research (LMS), which receives its core funding (intramural program) from the UK Medical Research Council. J.S. acknowledges support by the Center of Nanoscience (CeNS) of Ludwig-Maximilians-Universität as well as funding from the Deutsche Forschungsgemeinschaft (DFG) under grant STI673-2-1 and from the European Research Council under ERC grant agreement 758124. The Single Molecule Imaging Group is funded by a core grant of the MRC–London Institute of Medical Sciences (UKRI MC-A658-5TY10), a Wellcome Trust Collaborative Grant (P67153), and a BBSRC CASE-studentship (to M.D.N.). **Author contributions:** P.G.-E. expressed and purified yeast cohesin and Scc2-Scc4 proteins and performed biochemical assays. J.D. expressed and purified yeast condensin. I.A. expressed and purified human cohesin. P.G.-E., M.D.N., and A.L. collected optical tweezers datasets. P.G.-E., M.D.N., and J.S. processed optical tweezers data. J.H. and L.T. performed ATPase assays. R.A. prepared electron microscopy grids and collected and processed electron microscopy images. H.K. and A.M. performed mass spectrometry analysis. J.H. and J.S. performed, collected, and analyzed DNA curtain datasets. P.G.-E. and L.A. conceived the project. L.A. wrote the manuscript. L.A., D.S.R., and J.S revised the manuscript.

**Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The plasmid for the expression of human cohesin was a gift by the laboratory of J. M. Peters. Requests should be submitted to IMP Vienna.

Submitted 9 July 2019
Accepted 22 October 2019
Published 27 November 2019
10.1126/sciadv.aay6804

**Citation:** P. Gutierrez-Escribano, M. D. Newton, A. Llauró, J. Huber, L. Tanasie, J. Davy, I. Aly, R. Aramayo, A. Montoya, H. Kramer, J. Stigler, D. S. Rueda, L. Aragon, A conserved ATP- and Scc2/4-dependent activity for cohesin in tethering DNA molecules. *Sci. Adv.* **5**, eaay6804 (2019).

Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/11/eaay6804/DC1

# Supplementary Materials for

## A conserved ATP- and Scc2/4-dependent activity for cohesin in tethering DNA molecules

Pilar Gutierrez-Escribano, Matthew D. Newton, Aida Llauró, Jonas Huber, Loredana Tanasie, Joseph Davy, Isabel Aly, Ricardo Aramayo, Alex Montoya, Holger Kramer, Johannes Stigler*, David S. Rueda*, Luis Aragon*

*Corresponding author Email: stigler@genzentrum.lmu.de (J.S.); david.rueda@imperial.ac.uk (D.S.R.); luis.aragon@csc.mrc.ac.uk (L.A.)

Published 27 November 2019, *Sci. Adv.* **5**, eaay6804 (2019)
DOI: 10.1126/sciadv.aay6804

**The PDF file includes:**



**Other Supplementary Material for this manuscript includes the following:**

(available at advances.sciencemag.org/cgi/content/full/5/11/eaay6804/DC1)

Movie S3 (.mp4 format). Time-lapse videos showing cohesin tethering.
Movie S4 (.mp4 format). Time-lapse videos showing sliding of intermolecular bridges in a quadruple-trap optical tweezer.
Movie S5 (.mp4 format). Time-lapse videos showing sliding of intermolecular bridges in a quadruple-trap optical tweezer.
Movie S6 (.mp4 format). Time-lapse video showing pulling on intermolecular bridges in a quadruple-trap optical tweezer.

## Supplementary Figure 1

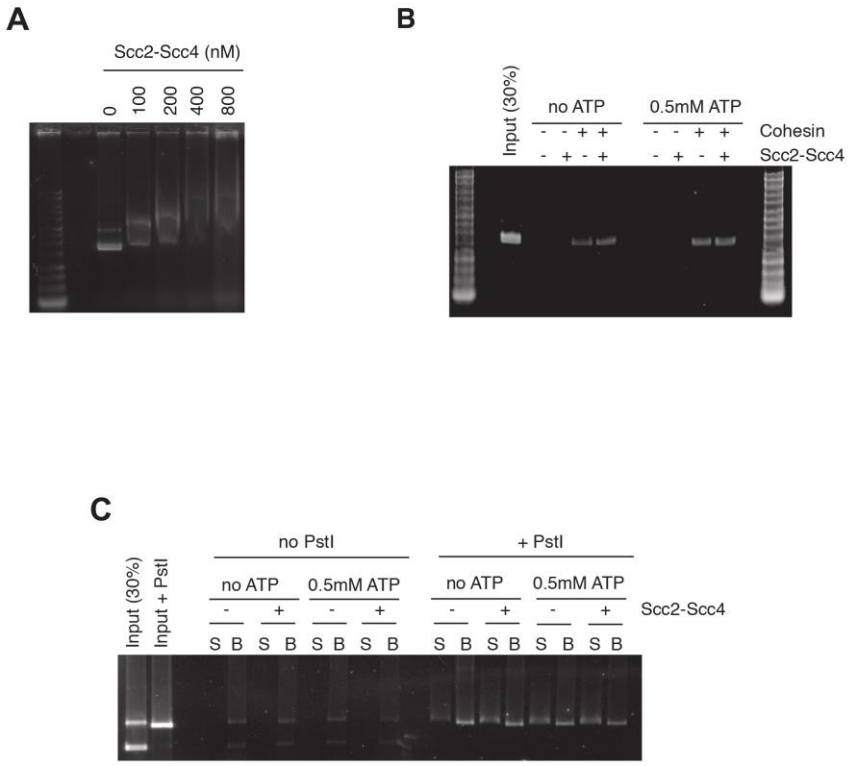



**Fig. S1. Topological loading of yeast cohesin on plasmid DNA. A**. Electrophoretic mobility shift assays using pUC19 as substrate and the indicated Scc2-Scc4 complex concentrations, showing that purified Scc2/Scc4 complex has DNA binding activity. **B.** Agarose gel electrophoresis showing recovered DNA after cohesin loading and immunoprecipitation with and without 0.5mM ATP both in the presence and absence of Scc2/4 complex. Topological assays were done as in (6). **C**. Gel image of recovered DNA in supernatant (S) and cohesin-bound bead (B) fractions after linearisation of immunoprecipitated cohesin-bound DNA by PstI digestion.

## Supplementary Figure 2

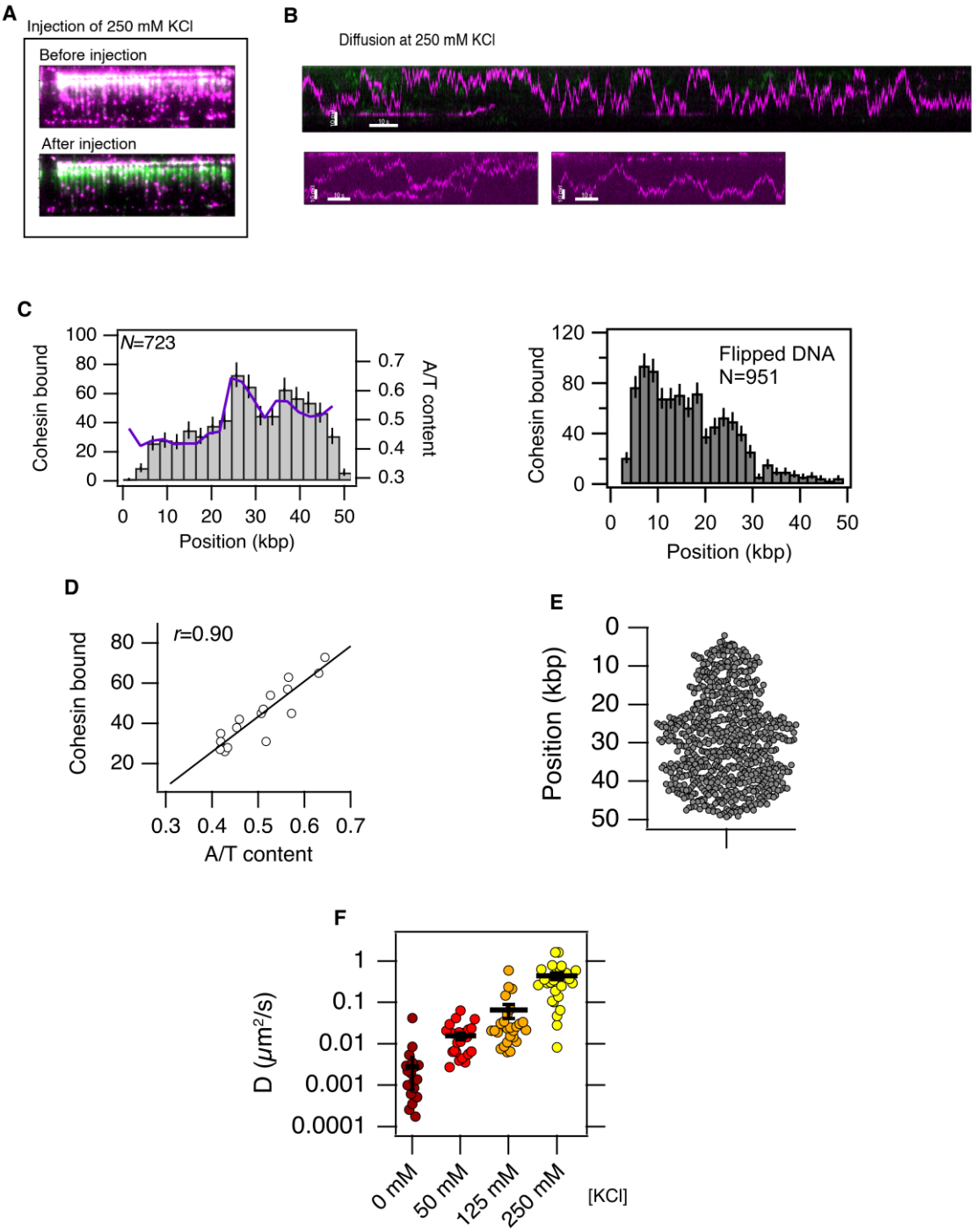


**Fig. S2. Analysis of yeast cohesin on DNA curtains. A**. Image of cohesin tagged with quantum dots (QD) (magenta) bound at 25 mM KCl to λ-DNA before injection of high ionic strength buffer (top). Image of cohesin (magenta) bound to λ-DNA (green) after injection of 250 mM KCl buffer (bottom). **A**. Representative kymograph illustrating cohesin diffusion at 250 mM KCl. **C**. Distribution of bound cohesins on λ-DNA. A/T content is indicated. Error bars: 68% confidence intervals. Distribution of bound cohesins on flipped λ-DNA is shown (right graph). **D**.Correlation between cohesin localisation and A/T nucleotide content on λ-DNA (Pearson's r = 0.90). **E**. Initial binding positions of individual cohesin to λ-DNA. **F**. Diffusion coefficients for cohesin movement in buffers of different of ionic concentration.

**Supplementary Figure 3**

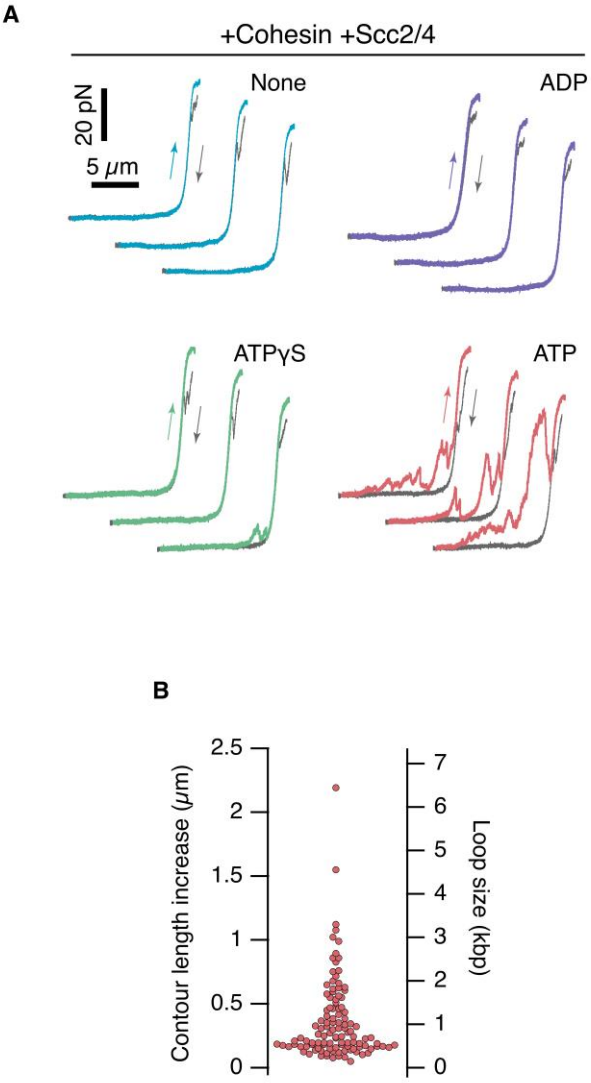


**Fig. S3. Intramolecular cohesin bridging requires ATP. A**. FE curves for λ-DNA pre-incubated with 1 nM cohesin, 2.5 nM Scc2-Scc4 complex and ATP analogues. After capturing a single DNA molecule between two optically trapped beads, DNA was incubated in the presence of protein in a relaxed conformation (3 μm bead distance) for 30s in 50mM NaCl and then moved to a buffer channel with 50mM NaCl for extension and measurements. Only ATP exhibits DNA bridging rupture events. Note that one of the FE curves (first curve in the ATP set) was used to illustrate bond-rupture events in Fig. 2A. **B**. Distribution of cohesin-induced DNA bridge loop sizes calculated from bond-rupture events in FE curves.

## Supplementary Figure 4

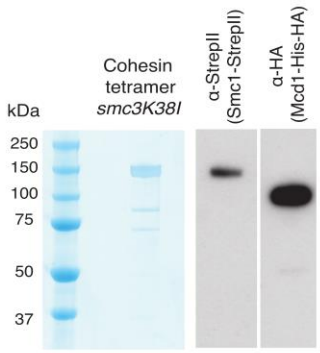



**Fig. S4. Purification of budding yeast Cohesin ATPase mutant.** Purified cohesin tetramer containing Smc1, Smc3-K38I, Mcd1 and Scc3 was analysed by SDS-PAGE electrophoresis followed by Coomassie Blue staining. Western analysis showing Smc1-strep and Mcd1 HA is included.

## Supplementary Figure 5

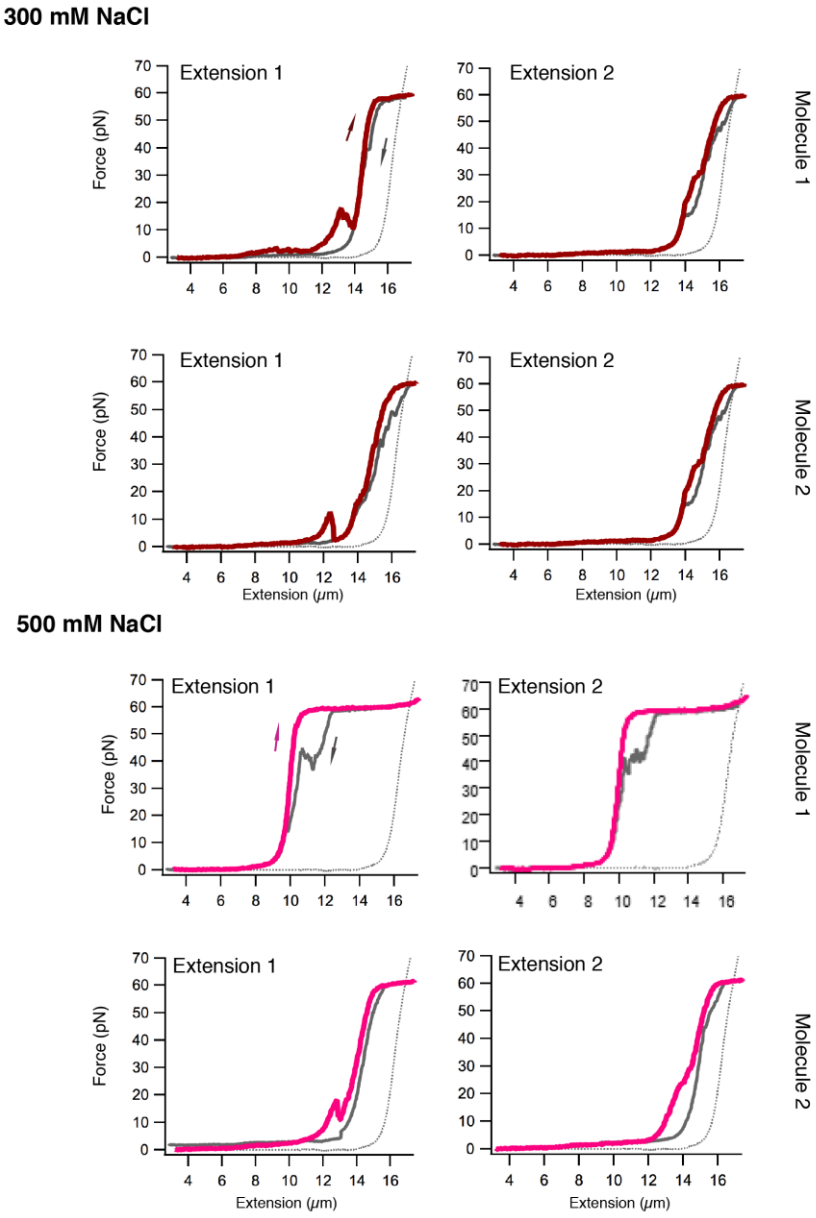


**Fig. S5. Permanent cohesin bridges are not displaced by physical stretching of λ-DNA.**
FE curves during sequential extension and relaxation cycles at 300mM and 500mM NaCl.
After capture of λ-DNA between the two optically trapped beads, DNA is relaxed (3μm bead distance) and incubated for 30 seconds in the protein channel (1 nM cohesin and 2.5 nM complex and 1 mM ATP in 50mM NaCl). DNA is moved to a buffer channel (either 300mM NaCl or 500mM NaCl as indicated) before re-extension to test for DNA bridges. After confirmation of the bridges and full extension of the molecule (FE curves on the left), the DNAs are relaxed in the same channel (either 300mM NaCl or 500mM NaCl as indicated) and re-extended for a second time. FE curves of the second re-extension (FE curves on the right) show that the DNA bridge has not been displaced. Two independent molecules are shown. The first extension is also shown in Fig. 3D-F.

**Supplementary Figure 6**

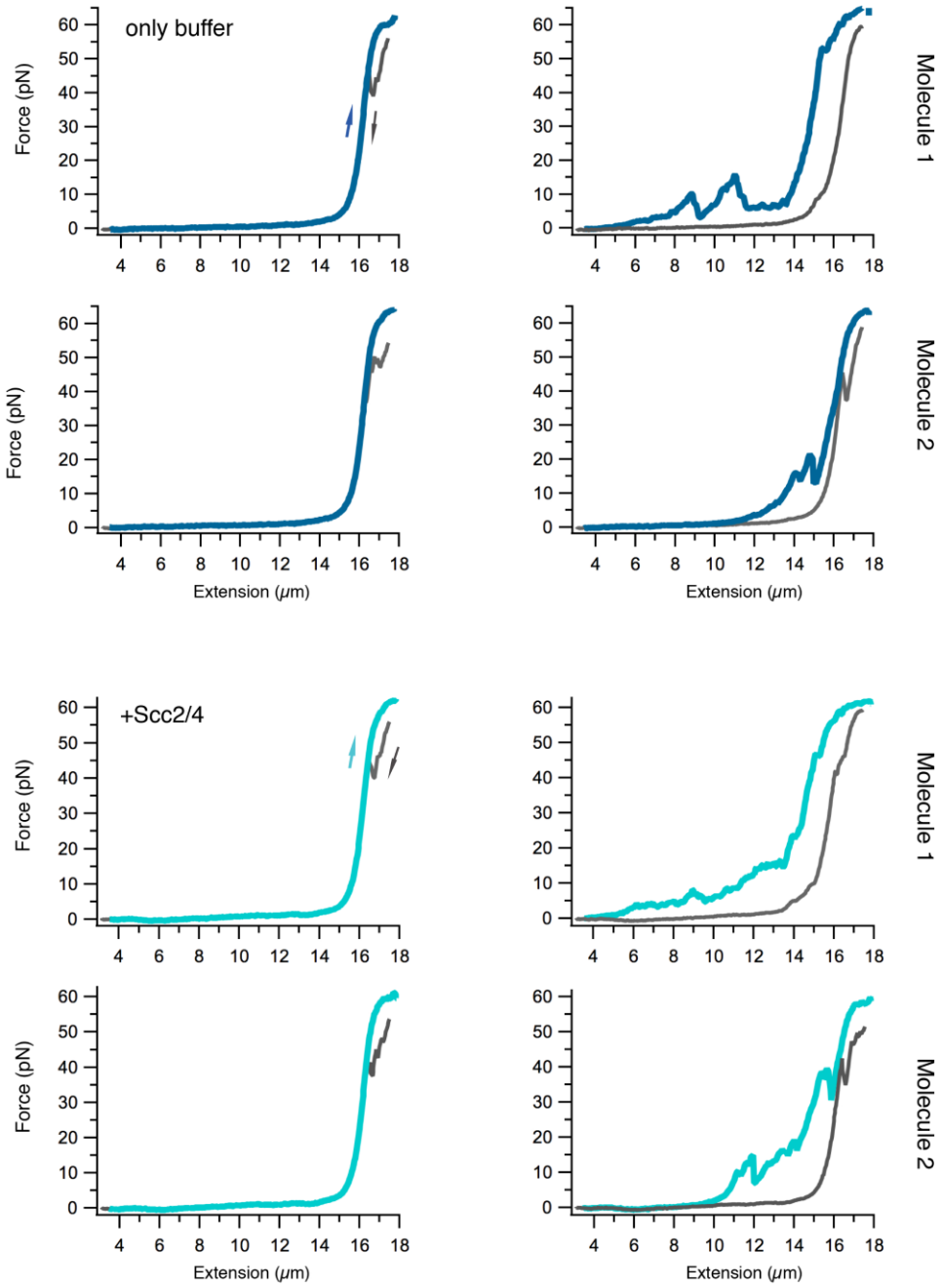


**Fig. S6. Cohesin does not capture two λ-DNAs in sequential steps.** λ-DNA incubated with 1nM cohesin, 2.5 nM Scc2-Scc4 complex and 1mM ATP in an extended conformation. The DNAs were moved to a buffer channel (50mM NaCl) in the presence of 1mM ATP (buffer only - dark blue- top two left FE curves) or 2.5nM Scc2-Scc4 complex and 1mM ATP (+Scc2/4 - light blue – bottom two left FE curves). The λ-DNA molecules were then incubated in a relaxed position (3μm bead distance) for 30s. DNAs were then moved to an only buffer channel (125mM NaCl containing 1mM ATP) and re-extended (right FE curves; top - only buffer - during the first relaxation, bottom buffer plus Scc2/4 during the first relaxation). The same molecules were then relaxed (3μm bead distance) and incubated (for 30s) with 1nM cohesin, 2.5 nM Scc2-Scc4 complex and 1mM ATP in 50mM NaCl. Finally, the DNA was moved to a different channel with 1mM ATP in 125mM NaCl and re-extended. FE curves of the final re-extension are shown (left FE curves; top- only buffer - during the first relaxation, -bottom- buffer plus Scc2/4 during the first relaxation). Only FE curves incubated with cohesin in a relaxed conformation show the presence of DNA bridging rupture events (i.e. FE curves on the right).

**Supplementary Figure 7**

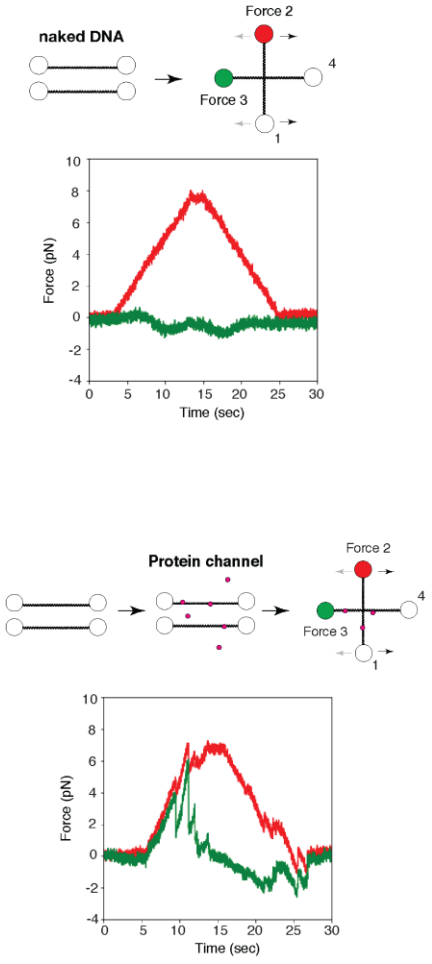
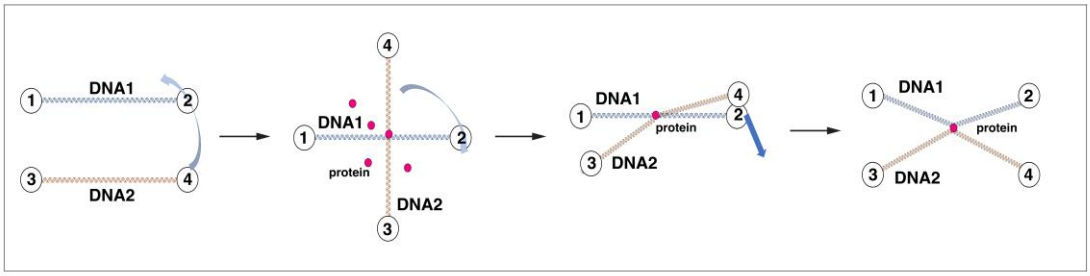


**Fig. S7. DNA friction experiments confirm the presence of cohesin complexes on extended λ-DNA.** Friction experiments were performed on a quadruple-trap optical tweezer system as described in (*11*). Two molecules of λ-DNA were tethered independently between two pairs of beads. Beads were then moved to the protein channel, containing 1nM cohesin and 2.5nM Scc2/4 complex in 50mM NaCl plus 1mM ATP, and incubated separated in an extended position (~14 μm bead distance). After incubation, the beads were moved to the buffer channel containing 125mM NaCl and 1mM ATP and crossed, so contact between both DNA molecules was established. Beads 1 and 2 were then moved simultaneously in the x axis to cause the sliding of one of the DNA molecules on top of the other. Forces in both beads 2 (sliding DNA molecule) and 3 (static DNA molecule) were recorded. A negative control with naked DNA is included (left panel). The presence of bound protein was identified by the appearance of abrupt changes in both force 2 and force 3 caused by the movement of beads 1 and 2 (right panel). For graphical representation, force data were downsampled to 100Hz.

## Supplementary Figure 8



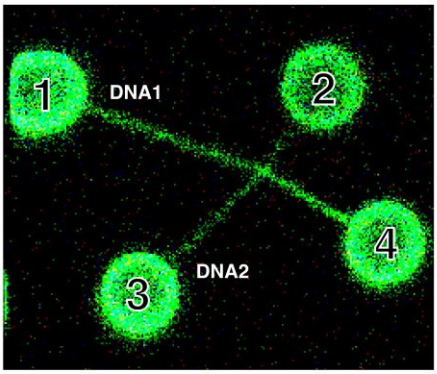




**Fig. S8. Generation of permanent cohesin bridges using a quadrupole-trap optical tweezer.** **A.** Schematic representation of the experimental design for the quadrupole trap optical tweezer to generate permanent cohesin bridges. First, a pair of λ-DNAs (DNA1 and DNA2) are trapped between two pairs of beads (beads 1 and 2 trap DNA1, while beads 3 and 4 trap DNA2) and kept extended (15μm bead distance). DNA2 is manipulated using beads 3 and 4, moved over DNA1 and positioned at a 90 angle, then relaxed (3μm bead distance). The crossed DNAs are moved to a channel containing 1 nM cohesin and 2.5 nM Scc2-Scc4 complex and 1mM ATP in 50mM NaCl and incubated for 30s. The crossed DNAs are moved to a channel containing 1mM ATP in 300mM NaCl and DNA2 is extended and moved back to its original position using beads 3 and 4. We used 50nM of SYTOX Orange to visualise the bridged DNA (right image). DNA molecules are in green, beads are numbered.

**Supplementary Figure 9**

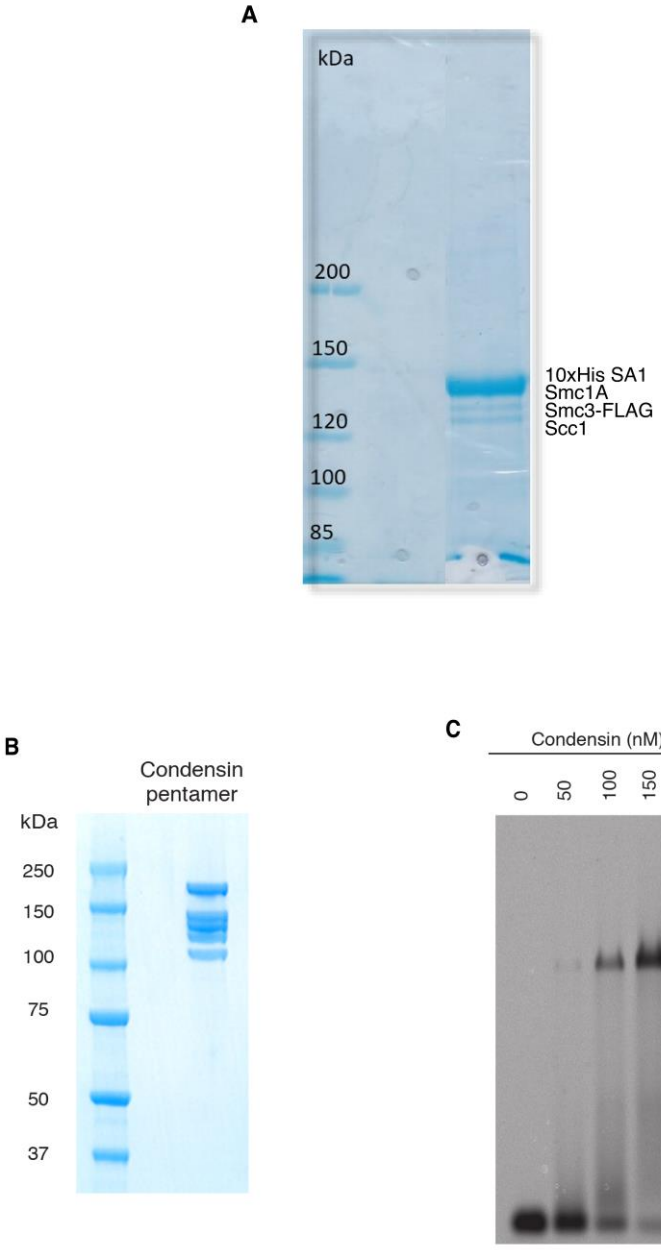


**Fig. S9. Purification of human Cohesin and yeast condensin. A**. Purified cohesin tetramer containing hSmc1A, Smc3-FLAG, Scc1 and 10xHis SA1 was analysed by SDS-PAGE electrophoresis followed by Coomassie Blue staining. Purification was done using biGBac vector (pBIG1c) as described previously (*3*). **B.** Purified condensin pentamers containing Smc2, Smc4, Brn1 and Ycs4 and Ycg1 was analysed by SDS-PAGE electrophoresis followed by Coomassie Blue staining. Purifications were done as in (*1, 2*). **C.** Electrophoretic mobility shift assays with a 6-carboxyfluorescein–labelled 41-bp dsDNA substrate (100nM) and the indicated protein concentrations.

**Supplementary Figure 10**

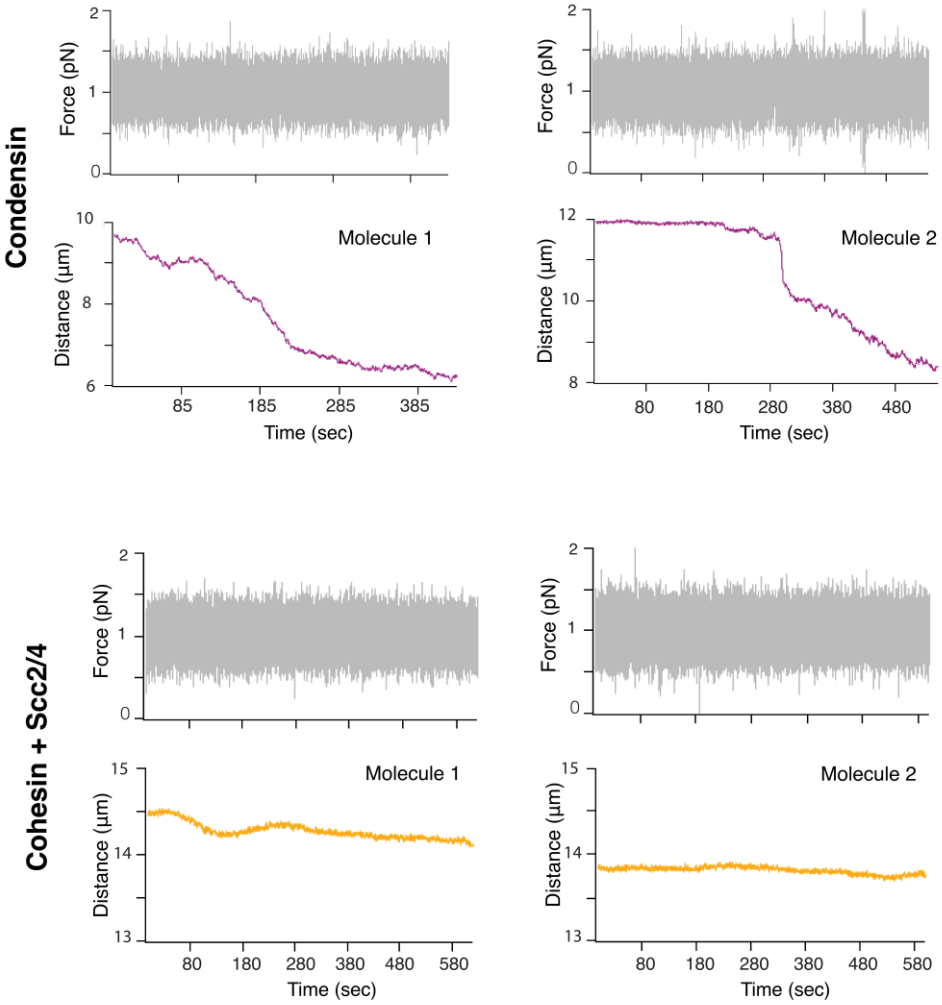


**Fig. S10. Budding yeast condensin, but not cohesin, compacts λ-DNA against 1 pN stretching force.** Examples of DNA compaction traces for λ-DNA molecule extended using a force of 1pN (top). The DNAs were tethered between two beads. One bead was clamped (fixed) while a 5pN force was applied to the second bead to maintain the molecule extended. The DNA was then incubated in the presence of 1nM condensin (1mM ATP in 50mM NaCl) (top- condensin - magenta traces) or 1nM Cohesin and 2.5 nM Scc2-Scc4 complex (1mM ATP in 50mM NaCl) (bottom -cohesin - - yellow traces). Extended DNAs were then moved to a different channel containing 1mM ATP in 50mM NaCl and the extension force was reduced to 1pN. The distance between the beads was recorded over time. Only condensin was able to reduce the distance between the beads over time consistent with a DNA compaction activity. Two independent molecule traces for each complex is shown. An additional trace is shown in Fig. 4F-H. For graphical representation, force data were downsampled to 100Hz.

Table S1. Mass spectrometry analysis of cohesin wild type and ATPase mutant (Smc3-K38I) tetramer complexes and the loader complex Scc2-Scc4.

Protein identifications for cohesin tetramer (Wt and K38I) and Scc2-Scc4 purifications.

Table S2. Mass spectrometry analysis of cohesin ATPase mutant (Smc3-K38I) tetramer peptides showing peptides containing the K38I mutation for SMC3.

Identifications of K38I peptides in purifications of cohesin ATPase mutant (Smc3-K38I).

Movies S1 to S3. Time-lapse videos showing cohesin tethering.

Time lapse videos showing cohesin tethering. Individual pairs of double-tethered DNA curtains (DNA molecules are in green) bound by cohesin is in magenta are shown.

Movies S4 and S5. Time-lapse videos showing sliding of intermolecular bridges in a quadruple-trap optical tweezer.

Time lapse videos showing sliding of intermolecular bridges in a quadruple-trap optical tweezer. Images are shown in Fig. 4C. DNA molecules are in green.

Movie S6. Time-lapse video showing pulling on intermolecular bridges in a quadruple-trap optical tweezer.

Time lapse video showing pulling on intermolecular bridges in a quadruple-trap optical tweezer. Images are shown in Fig. 4D. DNA molecules are in green.

## II 4. 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, Johannes Stigler (229)

**Summary:**

CTCF plays a fundamental role in genome architecture by acting as an anchor point for loop extruding cohesin at boundaries of topologically associating domains (TADs) (12, 18, 19). These domains control 3D genomic interactions between transcription regulatory domains, with interactions occurring more frequently within TADs than with neighbouring regions (16, 354). TAD boundary formation depends on a biochemical interaction of cohesin's RAD21 and SA subunits with CTCF's N-terminus (212, 213) and in some cases additionally requires CTCF clustering and RNA binding (168, 226, 253). However, an interaction between CTCF's C-terminus and SA has also been observed (217) and it has been shown that CTCF colocalised with SA *in vivo* independently of other cohesin subunits (216). The exact mechanism of CTCF-cohesin interaction and TAD boundary formation at CTCF-binding sites (CBSs), therefore, remains elusive.

Single-molecule fluorescence microscopy with parallelised λ-DNA strands containing multiple CBSs (DNA curtains) revealed that CTCF is enriched on CBSs by a higher lifetime and slower diffusion on CBSs compared to unspecific DNA. CTCF's lifetime was not influenced by the cooperativity of CTCFs on multiple close-by CBSs. Truncation mutants showed that enrichment depends on both the inner and outer zinc fingers (ZFs) but not on CTCF's termini. Cohesin's SA subunit was enriched on AT-rich regions by faster diffusion and a lower lifetime on GC-rich regions. If preincubated with CTCF, SA was localised at CBSs by a direct CTCF ZF-mediated interaction, increasing SA's lifetime and salt stability on DNA curtains. This suggests a recruitment mechanism of cohesin to TAD boundaries via its SA subunit since SA can diffuse on DNA until it is stabilised at positions of CBS-bound CTCF.

DNA-bound CTCF was able to additionally bind RNA, which stabilised CTCF even on unspecific sites. Photobleaching experiments revealed that this capture depends on CTCF oligomerisation. CTCF oligomers could therefore play a role in recruiting additional RNA-binding proteins like cohesin's SA subunit to CBSs, since SA was enriched on RNA structures on DNA curtains and evicted from curtains DNA by flushing in RNA, therefore having a higher affinity for RNA than DNA.

Finally, this study displayed that CTCF and CTCF-SA complexes are pushed off CBSs by transcribing T7 polymerases. Pushing CTCF only mildly influenced transcription speeds but led to an increased number of polymerase pauses and snapbacks, possibly via bridging of multiple DNAs/RNAs by CTCF. This shows that CTCF can directly influence transcription and also that transcriptional processes can relocate CTCF boundaries and thereby impact TAD formation.

We propose a mechanism for TAD boundary formation in which CTCF can stabilise SA on CBSs by the observed direct interaction with CTCF's ZF or possibly also indirectly via SA binding the RNA captured by CTCF oligomers. SA could then recruit additional cohesin subunits and CTCF's long lifetime on CBSs facilitates the stable anchoring of cohesin at TAD boundaries.

**Author contribution:**

I purified CTCF and SA variants and fluorescently labelled them. I performed CTCF and SA DNA curtains measurements and analysed enrichment data. I analysed the DNA curtains' lifetime data together with Johannes Stigler and supported Loredana Tanasie and Johannes Stigler in analysing diffusion data. Loredana Tanasie and I performed single-molecule T7 polymerase measurements. I analysed single-molecule transcription experiments and performed bulk transcription experiments. I performed RNA capture measurements and analysed photobleaching and enrichment data of CTCF-RNA clusters. Together with all other authors, I wrote the paper.

*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

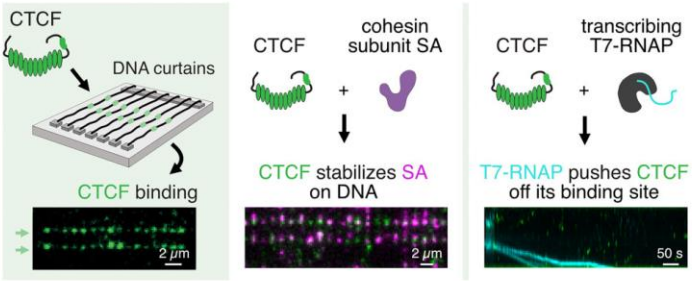
**Jonas Huber, Nicoleta-Loredana Tanasie, Sarah Zernia and Johannes Stigler***

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

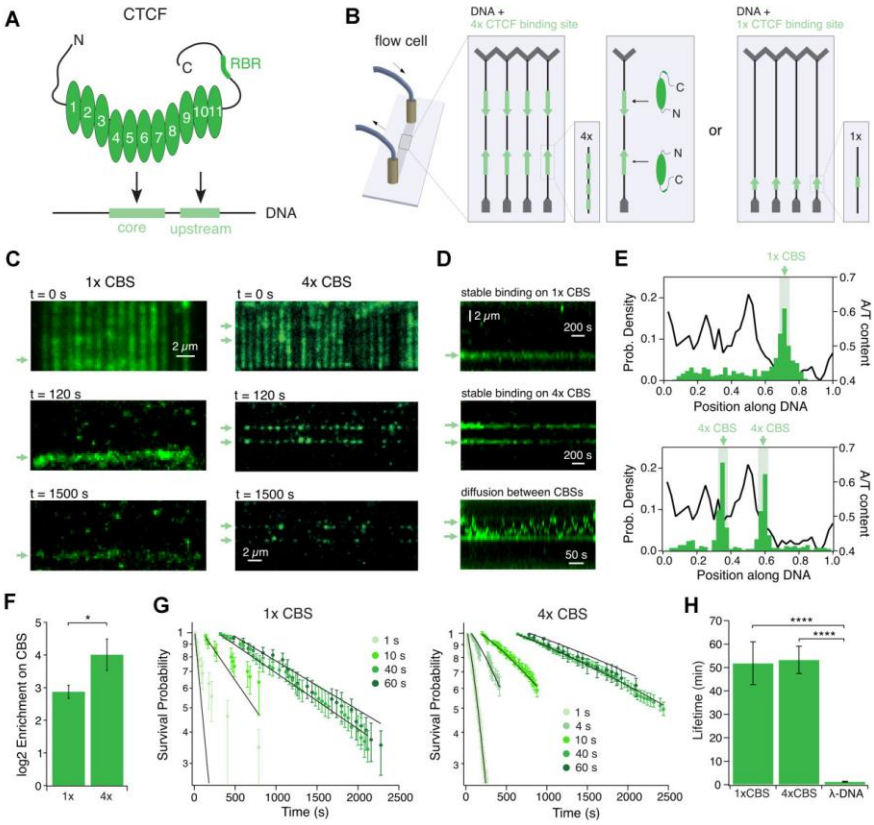



**Figure 1.** CTCF enriches on both, single and 4× CTCF-binding sites, with a lifetime of ∼50 min. (**A**) Schematic representation of CTCF containing the 11 ZFs, the RNA recognition motif (RBR), the elongated termini and the CTCF binding site (CBS) containing core and upstream motif. (**B**) Schematic 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 $\tau_v$ 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 $\tau_v$ 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 $\tau_v$ 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 $\tau$ 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

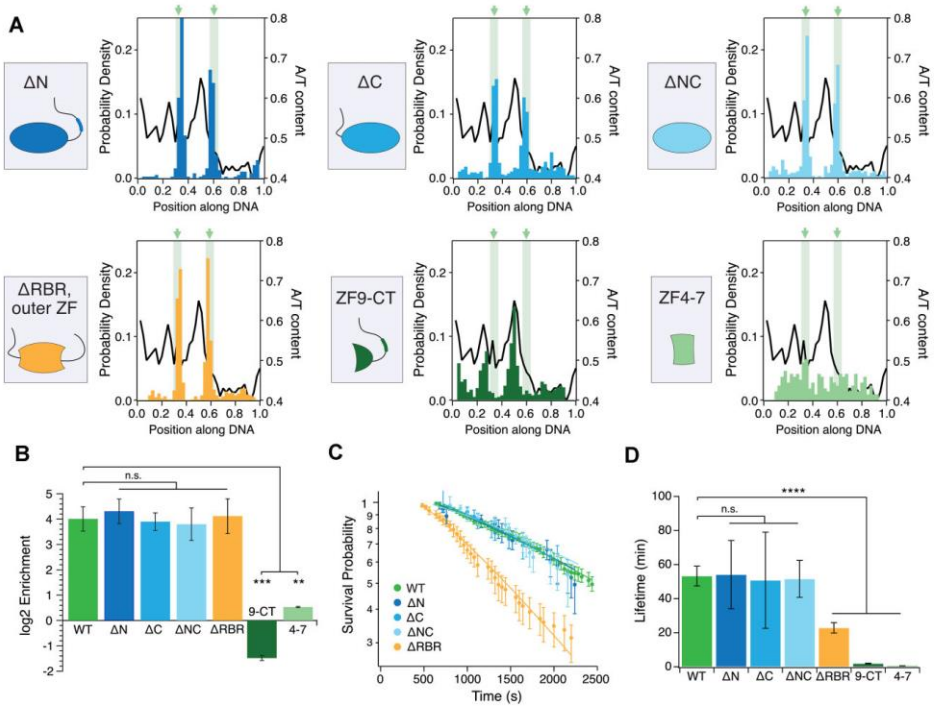



**Figure 2.** Binding behavior of CTCF variants on DNA curtains (**A**) Binding position of CTCF variants on 4× CTCF binding sites after 300 mM NaCl wash ($N = 390/666/396/651/343/88$ for ΔN/ΔC/ΔNC/ΔRBR/ZF9-CT/ZF4-7). ΔN/ΔC/ΔNC/ΔRBR labeled with AF568 and loaded at 10 nM concentration. ZF9-CT/ZF4-7 labeled 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-

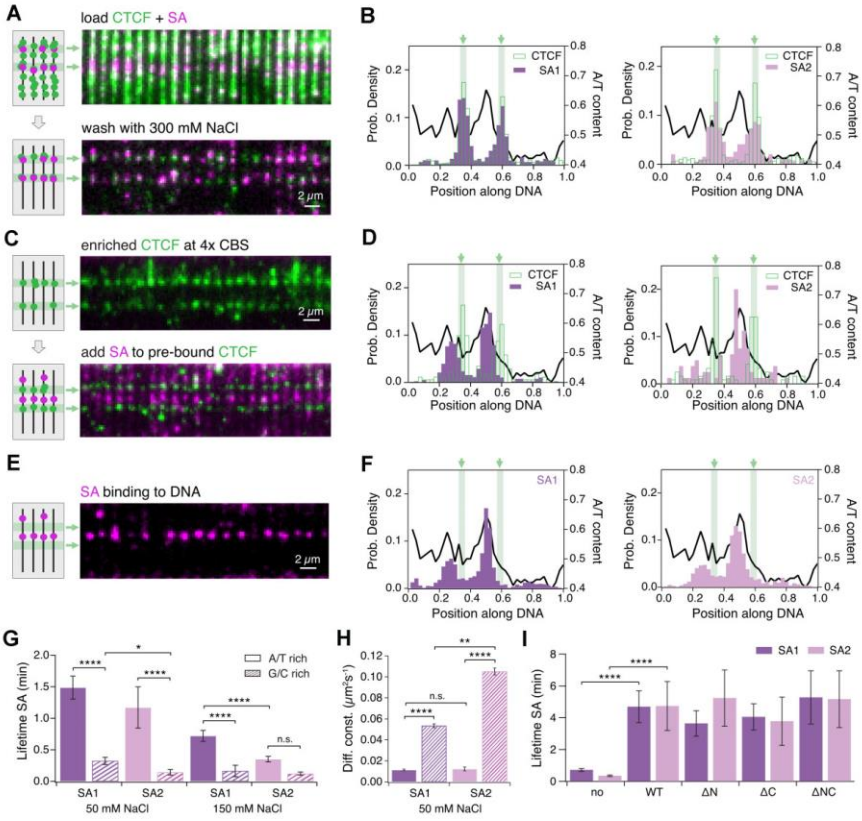



**Figure 3.** CTCF ZFs stabilize SA on CBSs (**A**) TIRF microscopy at 100 ms illumination time of fluorescently labeled CTCF (green) and SA (pink) preincubated (10 nM AF568-CTCF with 100 nM SA-LD655) before loading on DNA curtains, before (top) and after (bottom) 300 mM NaCl wash. (**B**) Histograms 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

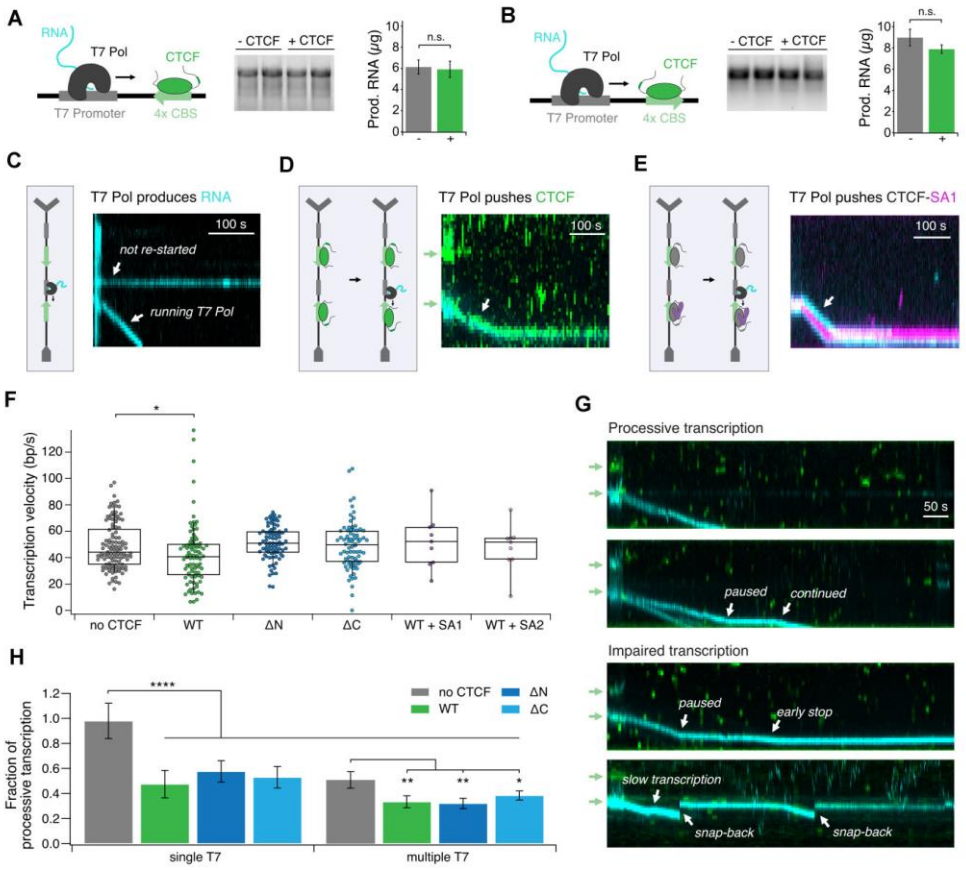



**Figure 4.** CTCF does not block transcription by T7 polymerase. (**A**) T7-Pol bulk transcription assay on a PCR product containing a T7-promoter as well as 4× CBSs with the downstream motif pointing towards the promoter. The length and amount of produced RNA was measured for three independent experiments 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

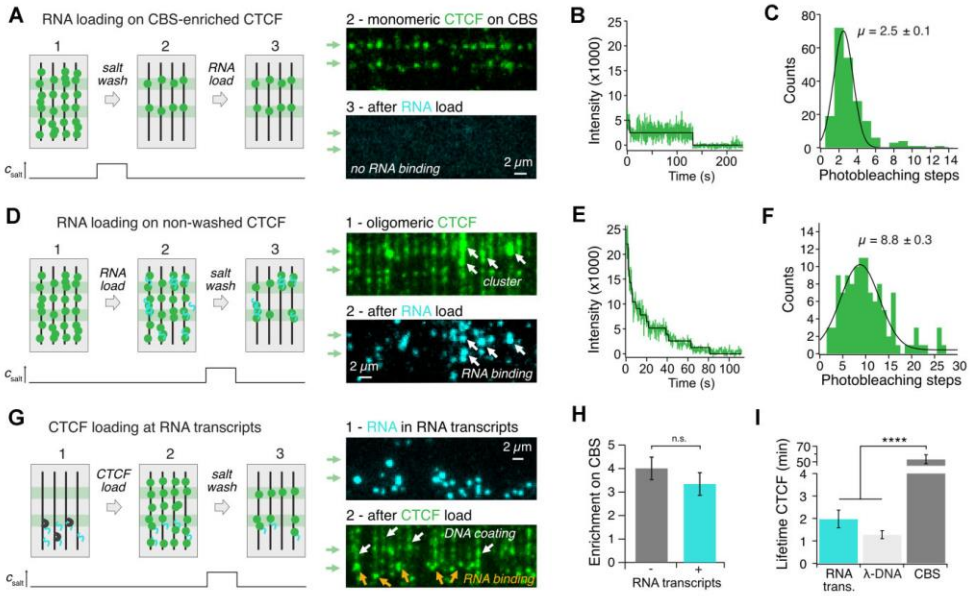



**Figure 5.** RNA recruitment by oligomeric CTCF. (**A**) 100 bp Cy3-UTP labeled RNA-loading (25 ng/μl) on CBS-enriched Alexa-Fluor 660-CTCF. Left: Experimental workflow. Right: TIRF microscopy at 100 ms illumination time of CBS-bound CTCF (green) before RNA load (top) and of Cy3-UTP labeled RNA (cyan) 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

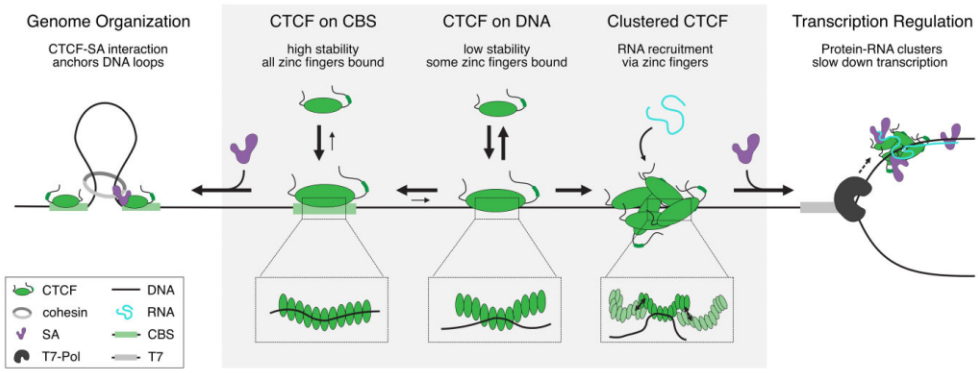



**Figure 6.** CTCF's nucleic acid interactions regulate diverse processes in the nucleus.

(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., Pękowska,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

Supplementary Information

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

Jonas Huber[1], Nicoleta-Loredana Tanasie[1], Sarah Zernia[1], Johannes Stigler[1,*]

[1] Gene Center Munich, Ludwig-Maximilians-Universität München, Munich, Germany
* To whom correspondence should be addressed: stigler@genzentrum.lmu.de

This file contains:

**Supplementary figures**

**Significance tests**

**Protein sequences**

## Supplementary figures

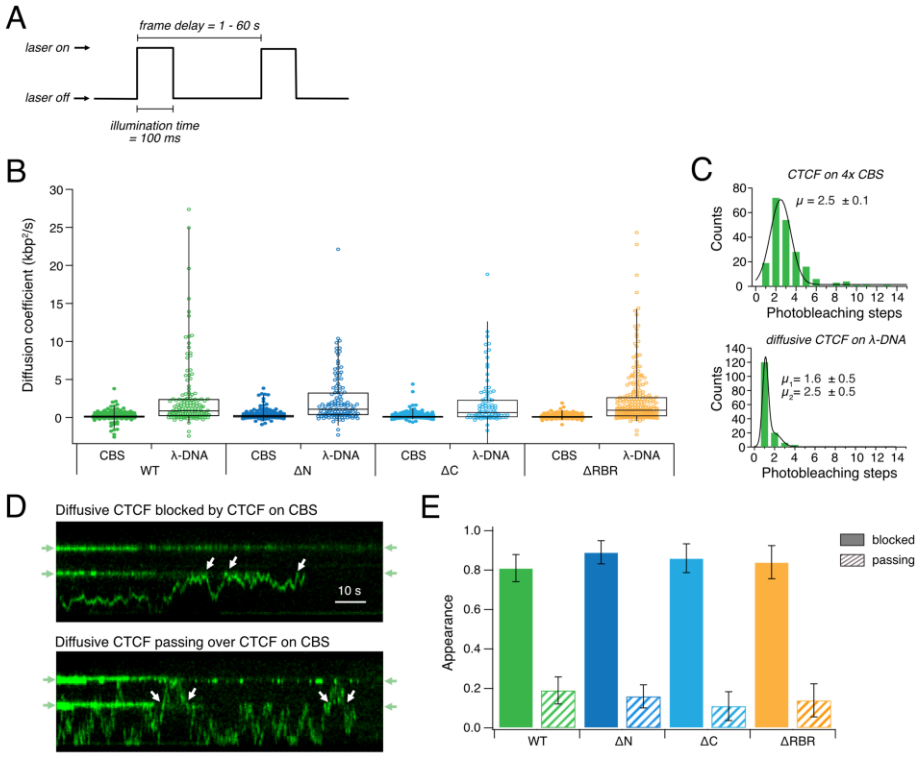



**Figure S1.** CTCF diffuses on non-CBS sites. **(A)** Scheme illustrating laser illumination during lifetime measurements. **(B)** Diffusion coefficients of AF568-CTCF WT and variants on 4x CBSs or λ-DNA at 10 nM concentration and 100 ms illumination time. For all variants, D is significantly higher on λ-DNA than CBSs. No significant difference in diffusive behavior between CTCF variants. **(C)** Photobleaching steps of non-diffusive (top) and diffusive (bottom) CTCF. **(D)** Representative kymograms showing diffusion behavior of WT CTCF. Top: White arrows indicate events where diffusive CTCF is blocked by CBS-bound CTCF. No recruitment of diffusive CTCF to the binding sites occurs. Bottom: White arrows indicate events where diffusive CTCF passes CBS-bound CTCF. **(E)** Quantification of blocking and passing events. No significant differences were observed between WT and CTCF variants.

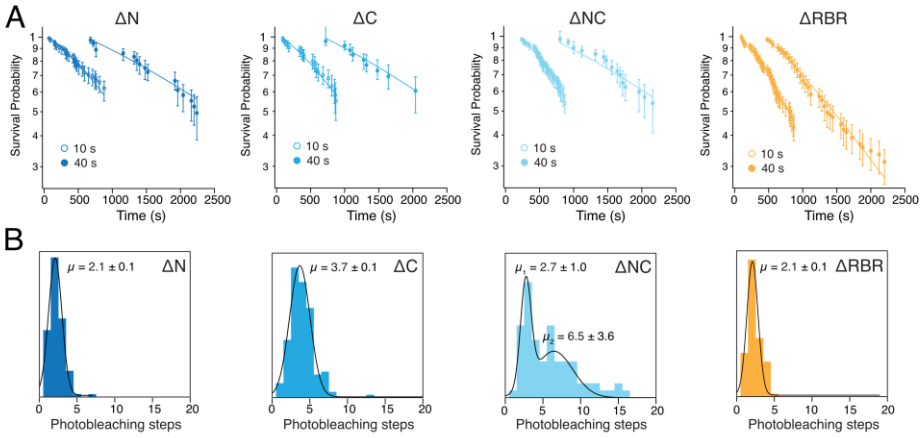



**Figure S2.** Lifetimes of AF568-CTCF variants at 10 nM concentration. **(A)** Lifetimes of CTCF variants at 10 and 40 s frame delay and 100 ms illumination time. **(B)** Photobleaching steps of CTCF variants binding to 4x CBSs. Black line: Multi-Gaussian fit.

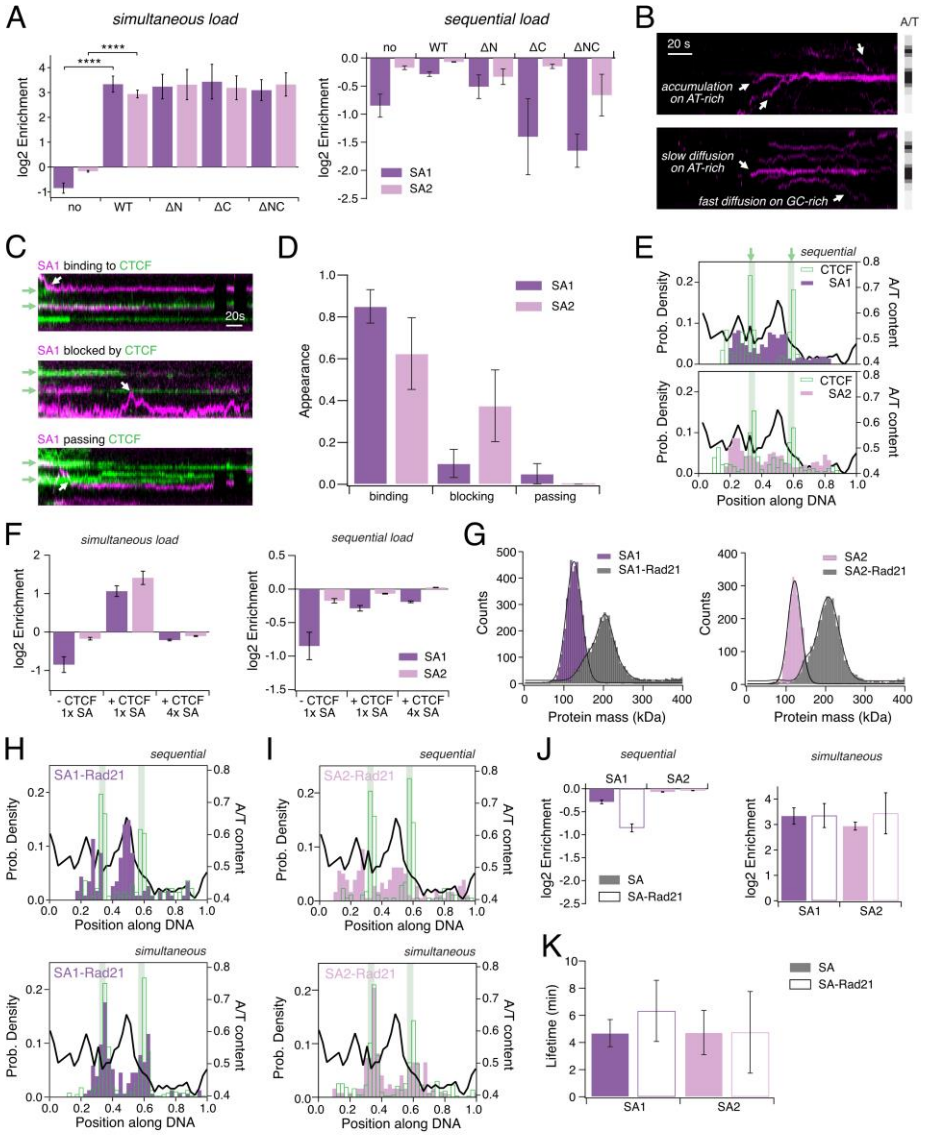



**Figure S3.** SA diffusion behavior and influence of Rad21 on CTCF-SA interaction. **(A)** Enrichment of SA1 (purple) and SA2 (pink) on CBSs. During simultaneous load (left), both SA-LD655s are enriched significantly more on CBSs when preincubated at 100 nM concentration with 10 nM AF568-CTCF WT compared to SAs loaded alone. SA is enriched similarly in presence of all CTCF variants. When 10 nM AF568-CTCF was bound first to DNA, followed by salt enrichment and by 100 nM SA-LD655 (sequential load, right), no significant SA enrichment at CBSs was observed with any CTCF variant. **(B)** Representative kymograms of SAs binding static to AT-rich and diffusing randomly on GC-rich

regions. **(C)** Representative kymograms of binding (top), blocking (middle) and passing (bottom) events observed for diffusive SAs (magenta) after sequential load on CTCF (green). **(D)** Fraction of binding, blocking and passing events observed for SA1 and SA2 after sequential load. No significant difference was found between SA1 and SA2. **(E)** Histograms of CTCF (green, top: N = 203; bottom: N = 410), SA1 (N = 580) and SA2 (N = 519) binding positions after sequential load at 4 x SA concentration (10 nM AF568-CTCF followed by 300 mM NaCl enrichment and 400 nM SA-LD655). CBSs are shown as green bars and the AT-ratio as a black line. **(F)** Left: enrichment of SAs on CBSs in absence of CTCF and at 1 x or 4 x SA concentrations, simultaneous load with CTCF, without salt enrichment. Right: enrichment of SAs on CBSs in absence of CTCF and at 1x or 4 x SA concentrations, sequential load after CTCF salt enrichment. **(G)** Mass photometry data of 148 kDa SA1 and 145 kDa SA2 in absence and presence (gray) of 59 kDa Rad21-MBP. **(H)** Histogram of SA1-Rad21 and CTCF binding positions for sequential load (top, CTCF, N = 137; SA1-Rad21, N = 114) and simultaneous load (bottom, CTCF, N = 397; SA1-Rad21, N = 153) experiments. **(I)** Same as (E) but using SA2-Rad21 (sequential, CTCF, N = 292; SA2-Rad21, N = 257; simultaneous, CTCF, N = 433; SA2-Rad21, N = 122). **(J)** Enrichment of SAs on CBSs in absence (solid) or presence (transparent) of Rad21 for sequential (left) and simultaneous load (right). Enrichments are independent from Rad21. **(K)** Lifetimes of SAs on CTCF in absence (solid) and presence (transparent) of Rad21. Lifetimes are independent of Rad21. All experiments were carried out at 100 ms illumination time.

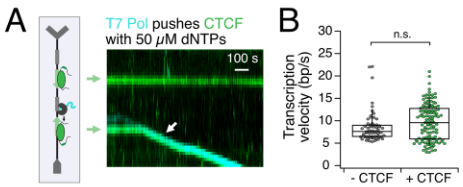



**Figure S4.** In vitro transcription assay at low nucleotide concentration **(A)** Illustration of in-vitro transcription assays and representative kymogram for T7-Pol pushing CTCF off its site during transcription at 50 $\mu$M nucleotide concentration (cyan = Cy3-UTP labeled RNA, green = AF568- CTCF) **(B)** Mean transcription velocities of T7-Pol alone and T7-Pol pushing CTCF WT at 50 $\mu$M nucleotide concentration.

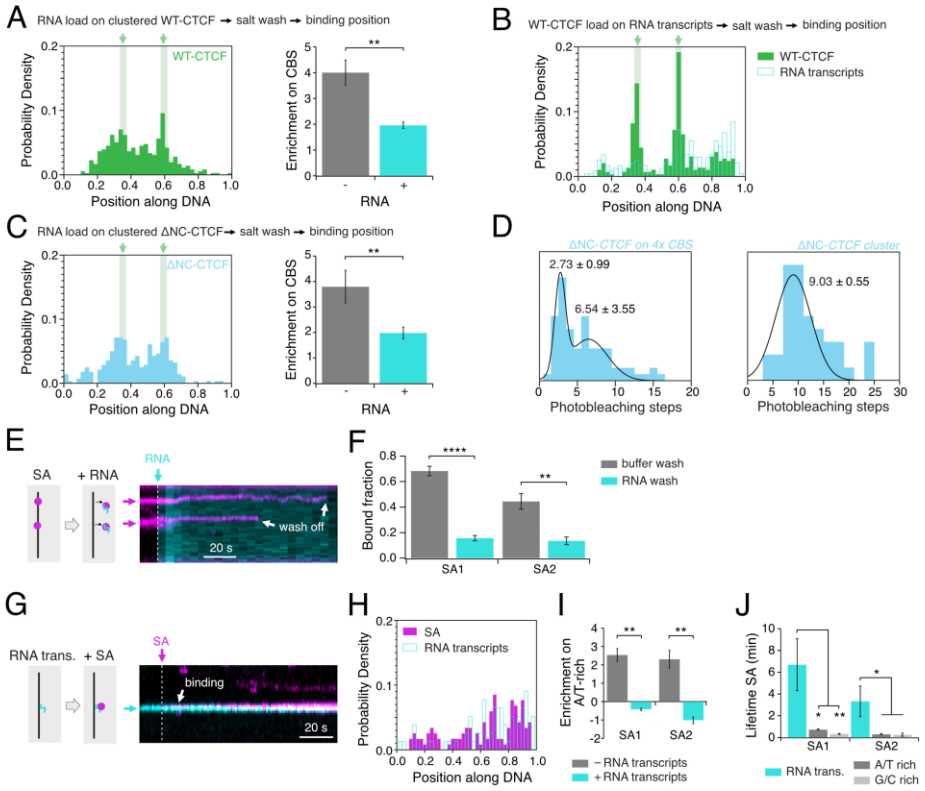



**Figure S5**. SA has a higher affinity for RNA than for DNA. **(A)** Left: Histogram of AF568-CTCF-WT-RNA cluster (N = 355) binding positions after salt enrichment (see Figure 5D). Right: CTCF WT-RNA clusters are significantly less enriched on CBSs than monomeric CTCF WT. **(B)** Histogram of CTCF WT binding positions after transcription and salt enrichment (see Figure 5G). CTCF preferentially binds to 4x CBSs but also colocalizes with RNA transcripts (CTCF, N = 292; RNA, N = 237). **(C)** Left: Histogram of AF568-CTCF-ΔNC-RNA cluster (N = 280) binding positions after salt enrichment. Right: ΔNC-RNA clusters are significantly less enriched on CBSs than monomeric ΔNC. **(D)** Histogram of photobleaching steps for ΔNC on 4x CBSs (same as Figure S2B-3) and in RNA-clusters. **(E)** Scheme and representative kymogram of SAs being washed off from λ-DNA by the addition of RNA to the DNA-curtain. **(F)** More 100 nM SA-LD655 is washed off in the presence of 25 ng/$\mu$l Cy3-UTP labeled RNA than in the presence of buffer. **(G)** Scheme and representative kymogram of SAs colocalizing with RNA transcripts. **(H)** Histogram of SA1 binding positions after transcription (SA1, N = 43; RNA, N = 155). **(I)** SA1 and SA2 are significantly less enriched on AT-rich regions in presence of RNA transcripts on the DNA. **(J)** SA1 and SA2 have a significantly higher lifetime on RNA transcripts than on AT-rich or GC-rich DNA regions. All experiments were carried out at 100 ms illumination time.

**Significance tests**

| | p-values | N |
|---|---|---|
| **Figure 1** | | |
| (F) 4x CBSs/1x CBS | 0.016 (t) | 427/477 |
| (H) 4x CBSs/1x CBS | 0.89 (z) | 701/201 |
| (H) 4x CBSs/λ-DNA | $< 10^{-6}$ (z) | 701/271 |
| (H) 1x CBS/λ-DNA | $< 10^{-6}$ (z) | 201/271 |
| **Figure S1** | | |
| (B) λ-DNA /CBS WT | $2.6 * 10^{-6}$ (t) | 148/261 |
| (B) λ-DNA /CBS ΔN | $< 10^{-6}$ (t) | 125/235 |
| (B) λ-DNA /CBS ΔC | $1.4 * 10^{-6}$ (t) | 89/288 |
| (B) λ-DNA / CBS ΔRBR | $< 10^{-6}$ (t) | 266/189 |
| (B) WT/ΔN λ-DNA | 0.37 (t) | 148/125 |
| (B) WT/ΔC λ-DNA | 0.19 (t) | 148/89 |
| (B) WT/ΔRBR λ-DNA | 0.24 (t) | 148/266 |
| (E) WT/ΔN | 0.48 (f) | 32/28 |
| (E) WT/ΔC | 0.72 (f) | 32/22 |
| (E) WT/ΔRBR | 1 (f) | 32/19 |
| **Figure 2** | | |
| (B) WT/ΔN | 0.22 (t) | 427/390 |
| (B) WT/ΔC | 0.63 (t) | 427/666 |
| (B) WT/ΔNC | 0.49 (t) | 427/396 |
| (B) WT/ΔRBR | 0.26 (t) | 427/651 |
| (B) WT/ZF9-CT | 0.0012 (t) | 427/343 |
| (B) WT/ZF4-7 | 0.0020 (t) | 427/88 |
| (D) WT/ΔN | 0.97 (z) | 701/125 |
| (D) WT/ΔC | 0.93 (z) | 701/80 |
| (D) WT/ΔNC | 0.89 (z) | 701/161 |
| (D) WT/ΔRBR | $3.4 * 10^{-6}$ (z) | 701/217 |
| (D) WT/ZF9-CT | $< 10^{-6}$ (z) | 701/155 |
| (D) WT/ZF4-7 | $< 10^{-6}$ (z) | 701/88 |
| **Figure 3** | | |
| (G) 50 mM NaCl SA1/SA2 GC | 0.012 (z) | 186/48 |
| (G) 50 mM NaCl SA1/SA2 AT | 0.37 (z) | 54/62 |
| (G) 50 mM NaCl GC/AT SA1 | $< 10^{-6}$ (z) | 54/186 |
| (G) 50 mM NaCl GC/AT SA2 | 0.0019 (z) | 62/48 |
| (G) 150 mM NaCl SA1/SA2 GC | 0.63 (z) | 8/20 |
| (G) 150 mM NaCl SA1/SA2 AT | $1.2 * 10^{-6}$ (z) | 65/62 |
| (G) 150 mM NaCl GC/AT SA1 | $< 10^{-6}$ (z) | 8/65 |
| (G) 150 mM NaCl GC/AT SA2 | 0.77 (z) | 20/62 |
| (H) SA1/SA2 GC | 0.0058 (z) | 537/510 |
| (H) SA1/SA2 AT | 0.92 (z) | 231/169 |
| (H) GC/AT SA1 | $2.1 * 10^{-5}$ (z) | 537/231 |
| (H) GC/AT SA2 | $< 10^{-6}$ (z) | 510/169 |

| | | |
|---|---|---|
| (I) WT/DNA SA1 | $7.7 * 10^{-5}$ (z) | 65/8 |
| (I) WT/DNA SA2 | 0.0044 (z) | 62/20 |
| (I) WT/ ΔN SA1 | 0.41 (z) | 65/28 |
| (I) WT/ΔC SA1 | 0.62 (z) | 65/20 |
| (I) WT/ΔNC SA1 | 0.76 (z) | 65/9 |
| (I) WT/ΔN SA2 | 0.83 (z) | 62/28 |
| (I) WT/ΔC SA2 | 0.66 (z) | 62/17 |
| (I) WT/ΔNC SA2 | 0.85 (z) | 62/10 |
| **Figure S3** | | |
| (A)WT/DNA SA1 simultaneous | $2.5 * 10^{-4}$ (t) | 332/1321 |
| (A)WT/DNA SA2 simultaneous | $3.3 * 10^{-6}$ (t) | 224/875 |
| (A)WT/ΔN SA1 simultaneous | 0.74 (t) | 332/197 |
| (A)WT/ΔN SA2 simultaneous | 0.31 (t) | 224/153 |
| (A)WT/ΔC SA1 simultaneous | 0.79 (t) | 332/251 |
| (A)WT/ΔC SA2 simultaneous | 0.41 (t) | 224/205 |
| (A)WT/ΔNC SA1 simultaneous | 0.39 (t) | 332/116 |
| (A)WT/ΔNC SA2 simultaneous | 0.22 (t) | 224/40 |
| (A)WT/DNA SA1 sequential | 0.22 (t) | 380/1321 |
| (A)WT/DNA SA2 sequential | 0.76 (t) | 80/875 |
| (A)WT/ΔN SA1 sequential | 0.73 (t) | 380/114 |
| (A)WT/ΔN SA2 sequential | 0.85 (t) | 80/88 |
| (A)WT/ΔC SA1 sequential | 0.55 (t) | 380/265 |
| (A)WT/ΔC SA2 sequential | 0.97 (t) | 80/242 |
| (A)WT/ΔNC SA1 sequential | 0.18 (t) | 380/207 |
| (A)WT/ΔNC SA2 sequential | 0.63 (t) | 80/95 |
| (D)SA1/SA2 binding | 0.31 (f) | 20/8 |
| (D)SA1/SA2 blocking | 0.12 (f) | 20/8 |
| (D)SA1/SA2 passing | 1.0 (f) | 20/8 |
| (F)WT/DNA 1 x SA1 simultaneous | 0.029 (t) | 531/1321 |
| (F)WT/DNA 1 x SA2 simultaneous | 0.029 (t) | 177/875 |
| (F)WT/DNA 4 x SA1 simultaneous | 0.13 (t) | 418/1321 |
| (F)WT/DNA 4 x SA2 simultaneous | 0.86 (t) | 1156/875 |
| (F)WT/DNA SA1 sequential | 0.22 (t) | 380/1321 |
| (F)WT/DNA SA2 sequential | 0.76 (t) | 80/875 |
| (F)WT/DNA 4 x SA1 sequential | 0.085 (t) | 580/1321 |
| (F)WT/DNA 4 x SA2 sequential | 0.59 (t) | 519/875 |
| (J)SA1/SA1Rad21 on WT simultaneous | 0.95 (t) | 332/137 |
| (J)SA2/SA2Rad21 on WT simultaneous | 0.31 (t) | 224/277 |
| (J)SA1/SA1Rad21 on WT sequential | 0.15 (t) | 380/153 |
| (J)SA2/SA2Rad21 on WT sequential | 0.90 (t) | 80/122 |
| (K)SA1/SA1Rad21 on WT | 0.51 (z) | 65/16 |
| (K)SA2/SA2Rad21 on WT | 1.0 (z) | 62/9 |
| **Figure 4** | | |
| (A)CTCF/no | 0.78 (t) | 3/3 |
| (B)CTCF/no | 0.18 (t) | 3/3 |

| | | |
|---|---|---|
| (F)T7/WT | 0.034 (t) | 111/88 |
| (F)T7/ΔN | 0.22 (t) | 111/84 |
| (F)T7/ΔC | 0.89 (t) | 111/76 |
| (F)T7/WT-SA1 | 0.63 (t) | 111/9 |
| (F)T7/WT-SA2 | 0.85 (t) | 111/9 |
| (H)T7single/T7multiple | $< 10^{-6}$ (f) | 49/118 |
| (H)T7single/WTsingle | $< 10^{-6}$ (f) | 49/40 |
| (H)T7single/WTmultiple | $< 10^{-6}$ (f) | 49/147 |
| (H)T7single/ΔNsingle | $4.1 * 10^{-6}$ (f) | 49/33 |
| (H)T7single/ΔNmultiple | $< 10^{-6}$ (f) | 49/134 |
| (H)T7single/ΔCsingle | $< 10^{-6}$ (f) | 49/34 |
| (H)T7single/ΔCmultiple | $< 10^{-6}$ (f) | 49/177 |
| (H)T7multiple/WTmultiple | 0.0056 (f) | 118/147 |
| (H)T7multiple/ΔNmultiple | 0.0031 (f) | 118/134 |
| (H)T7multiple/ΔCmultiple | 0.042 (f) | 118/177 |
| **Figure S4** | | |
| (B) T7/WT | 0.075 (t) | 61/121 |
| **Figure 5** | | |
| (H)RNA/noRNA enrichment 4x CBSs | 0.060 (t) | 292/427 |
| (I) RNA/4x CBSs | $< 10^{-6}$ (z) | 118/201 |
| (I) RNA/λ-DNA | 0.11 (z) | 118/274 |
| **Figure S5** | | |
| (A)RNAcluster/monomers WT | 0.0021 (t) | 355/427 |
| (C)RNAcluster/monomers ΔNC | 0.0022 (t) | 280/396 |
| (F)RNA/noRNA wash off SA1 | $< 10^{-6}$ (t) | 560/479 |
| (F)RNA/noRNA wash off SA2 | 0.00267 (t) | 148/198 |
| (I)RNA/noRNA AT-enrichment SA1 | 0.0029 (t) | 1321/43 |
| (I)RNA/noRNA AT-enrichment SA2 | 0.0088 (t) | 875/117 |
| (J)RNA/AT | 0.012 (z) | 420/260 |
| (J)RNA/AT | 0.032 (z) | 295/119 |
| (J)RNA/GC | 0.0075 (z) | 420/93 |
| (J)RNA/GC | 0.030 (z) | 295/146 |

**Table S1.** Number of molecules and p-values for all relevant experiments. Two tailed t-test: (t), Two tailed z-test: (z), Fisher's exact test: (f).

**Protein Sequences**

CTCF WT
6xHis – Halo – TEV site – CTCF WT – Flag

MGSSHHHHHHSSGTSLYKKAGLMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHG
NPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHD
WGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEG
TLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPV
PKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGEPT
TEDLYFQSDNTTLYTKVVMEGDAVEAIVEESETFIKGKERKTYQRRREGGQEEDACHLPQNQTD
GGEVVQDVNSSVQMVMMEQLDPTLLQMKTEVMEGTVAPEAEAAVDDTQIITLQVVNMEEQPINI
GELQLVQVPVPVTVPVATTSVEELQGAYENEVSKEGLAESEPMICHTLPLPEGFQVVKVGANGE
VETLEQGELPPQEDPSWQKDPDYQPPAKKTKKTKKSKLRYTEEGKDVDVSVYDFEEEQQEGLL
SEVNAEKVVGNMKPPKPTKIKKKGVKKTFQCELCSYTCPRRSNLDRHMKSHTDERPHKCHLCG
RAFRTVTLLRNHLNTHTGTRPHKCPDCDMAFVTSGELVRHRRYKHTHEKPFKCSMCDYASVEV
SKLKRHIRSHTGERPFQCSLCSYASRDTYKLKRHMRTHSGEKPYECYICHARFTQSGTMKMHIL
QKHTENVAKFHCPHCDTVIARKSDLGVHLRKQHSYIEQGKKCRYCDAVFHERYALIQHQKSHKN
EKRFKCDQCDYACRQERHMIMHKRTHTGEKPYACSHCDKTFRQKQLLDMHFKRYHDPNFVPA
AFVCSKCGKTFTRRNTMARHADNCAGPDGVEGENGGETKKSKRGRKRKMRSKKEDSSDSEN
AEPDLDDNEDEEEPAVEIEPEPEPQPVTPAPPPAKKRRGRPPGRTNQPKQNQPTAIIQVEDQNT
GAIENIIVEVKKEPDAEPAEGEEEEAQPAATDAPNGDLTPEMILSMMDRDYKDDDDK

CTCF ΔN
6xHis – Halo – TEV site – CTCF (Δ1-265) – Flag

MGSSHHHHHHSSGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWR
NIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFH
WAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVR
PLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTP
GVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGEPTTEDLYFQSG
SFQCELCSYTCPRRSNLDRHMKSHTDERPHKCHLCGRAFRTVTLLRNHLNTHTGTRPHKCPDC
DMAFVTSGELVRHRRYKHTHEKPFKCSMCDYASVEVSKLKRHIRSHTGERPFQCSLCSYASRD
TYKLKRHMRTHSGEKPYECYICHARFTQSGTMKMHILQKHTENVAKFHCPHCDTVIARKSDLGV
HLRKQHSYIEQGKKCRYCDAVFHERYALIQHQKSHKNEKRFKCDQCDYACRQERHMIMHKRTH
TGEKPYACSHCDKTFRQKQLLDMHFKRYHDPNFVPAAFVCSKCGKTFTRRNTMARHADNCAG
PDGVEGENGGETKKSKRGRKRKMRSKKEDSSDSENAEPDLDDNEDEEEPAVEIEPEPEPQPV
TPAPPPAKKRRGRPPGRTNQPKQNQPTAIIQVEDQNTGAIENIIVEVKKEPDAEPAEGEEEEAQP
AATDAPNGDLTPEMILSMMDRDYKDDDDK

CTCF ΔC
6xHis – Halo – TEV site – CTCF (Δ580-727) – Flag

MGSSHHHHHHSSGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWR
NIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFH
WAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVR
PLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTP
GVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGEPTTEDLYFQSD
NTTLYTKVVMEGDAVEAIVEESETFIKGKERKTYQRRREGGQEEDACHLPQNQTDGGEVVQDV
NSSVQMVMMEQLDPTLLQMKTEVMEGTVAPEAEAAVDDTQIITLQVVNMEEQPINIGELQLVQV
PVPVTVPVATTSVEELQGAYENEVSKEGLAESEPMICHTLPLPEGFQVVKVGANGEVETLEQGE

LPPQEDPSWQKDPDYQPPAKKTKKTKKSKLRYTEEGKDVDVSVYDFEEEQQEGLLSEVNAEKV
VGNMKPPKPTKIKKKGVKKTFQCELCSYTCPRRSNLDRHMKSHTDERPHKCHLCGRAFRTVTL
LRNHLNTHTGTRPHKCPDCDMAFVTSGELVRHRRYKHTHEKPFKCSMCDYASVEVSKLKRHIR
SHTGERPFQCSLCSYASRDTYKLKRHMRTHSGEKPYECYICHARFTQSGTMKMHILQKHTENV
AKFHCPHCDTVIARKSDLGVHLRKQHSYIEQGKKCRYCDAVFHERYALIQHQKSHKNEKRFKCD
QCDYACRQERHMIMHKRTHTGEKPYACSHCDKTFRQKQLLDMHFKRYHDPNFVPAAFVCSKC
GKTFTRRNTMARHADNCAGDYKDDDDK

CTCF ΔNC
6xHis – Halo – TEV site – CTCF (Δ1-265; Δ580-727) – Flag

MGSSHHHHHHSSGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWR
NIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFH
WAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVR
PLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTP
GVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGEPTTEDLYFQSG
SFQCELCSYTCPRRSNLDRHMKSHTDERPHKCHLCGRAFRTVTLLRNHLNTHTGTRPHKCPDC
DMAFVTSGELVRHRRYKHTHEKPFKCSMCDYASVEVSKLKRHIRSHTGERPFQCSLCSYASRD
TYKLKRHMRTHSGEKPYECYICHARFTQSGTMKMHILQKHTENVAKFHCPHCDTVIARKSDLGV
HLRKQHSYIEQGKKCRYCDAVFHERYALIQHQKSHKNEKRFKCDQCDYACRQERHMIMHKRTH
TGEKPYACSHCDKTFRQKQLLDMHFKRYHDPNFVPAAFVCSKCGKTFTRRNTMARHADNCAG
DYKDDDDK

CTCF ΔRBR
6xHis – Halo – TEV site – CTCF (Δ264-291; Δ521-614) – Flag

MGSSHHHHHHSSGTSLYKKAGLMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHG
NPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHD
WGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEG
TLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPV
PKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGEPT
TEDLYFQSDNTTLYTKVVMEGDAVEAIVEESETFIKGKERKTYQRRREGGQEEDACHLPQNQTD
GGEVVQDVNSSVQMVMMEQLDPTLLQMKTEVMEGTVAPEAEAAVDDTQIITLQVVNMEEQPINI
GELQLVQVPVPVTVPVATTSVEELQGAYENEVSKEGLAESEPMICHTLPLPEGFQVVKVGANGE
VETLEQGELPPQEDPSWQKDPDYQPPAKKTKKTKKSKLRYTEEGKDVDVSVYDFEEEQQEGLL
SEVNAEKVVGNMKPPKPTKIKKKGVKRPHKCHLCGRAFRTVTLLRNHLNTHTGTRPHKCPDCD
MAFVTSGELVRHRRYKHTHEKPFKCSMCDYASVEVSKLKRHIRSHTGERPFQCSLCSYASRDT
YKLKRHMRTHSGEKPYECYICHARFTQSGTMKMHILQKHTENVAKFHCPHCDTVIARKSDLGVH
LRKQHSYIEQGKKCRYCDAVFHERYALIQHQKSHKNEKRFKCDQCDYACRQERHMIMHKRTHT
GEAEPDLDDNEDEEEPAVEIEPEPEPQPVTPAPPPAKKRRGRPPGRTNQPKQNQPTAIIQVEDQ
NTGAIENIIVEVKKEPDAEPAEGEEEEAQPAATDAPNGDLTPEMILSMMDRDYKDDDDK

ZF4-7
6xHis – CTCF (Δ1-350; Δ461-727) – Flag

MGSSHHHHHHSSGFKCSMCDYASVEVSKLKRHIRSHTGERPFQCSLCSYASRDTYKLKRHMR
THSGEKPYECYICHARFTQSGTMKMHILQKHTENVAKFHCPHCDTVIARKSDLGVHLRKQHDYK
DDDDK

ZF9-CT
6xHis – CTCF (Δ1-489) – Flag

MGSSHHHHHHSSGLVPRGSHMKNEKRFKCDQCDYACRQERHMIMHKRTHTGEKPYACSHCD
KTFRQKQLLDMHFKRYHDPNFVPAAFVCSKCGKTFTRRNTMARHADNCAGPDGVEGENGGET
KKSKRGRKRKMRSKKEDSSDSENAEPDLDDNEDEEEPAVEIEPEPEPQPVTPAPPPAKKRRGR
PPGRTNQPKQNQPTAIIQVEDQNTGAIENIIVEVKKEPDAEPAEGEEEEAQPAATDAPNGDLTPE
MILSMMDRDYKDDDDK

SA1

10xHis – SA1 WT – S6

MHHHHHHHHHHSGGSMITSELPVLQDSTNETTAHSDAGSELEETEVKGKRKRGRPGRPPSTN
KKPRKSPGEKSRIEAGIRGAGRGRANGHPQQNGEGEPVTLFEVVKLGKSAMQSVVDDWIESYK
QDRDIALLDLINFFIQCSGCRGTVRIEMFRNMQNAEIIRKMTEEFDEDSGDYPLTMPGPQWKKFR
SNFCEFIGVLIRQCQYSIIYDEYMMDTVISLLTGLSDSQVRAFRHTSTLAAMKLMTALVNVALNLSI
HQDNTQRQYEAERNKMIGKRANERLELLLQKRKELQENQDEIENMMNSIFKGIFVHRYRDAIAEI
RAICIEEIGVWMKMYSDAFLNDSYLKYVGWTLHDRQGEVRLKCLKALQSLYTNRELFPKLELFTN
RFKDRIVSMTLDKEYDVAVEAIRLVTLILHGSEEALSNEDCENVYHLVYSAHRPVAVAAGEFLHK
KLFSRHDPQAEEALAKRRGRNSPNGNLIRMLVLFFLESELHEHAAYLVDSLWESSQELLKDWEC
MTELLLEEPVQGEEAMSDRQESALIELMVCTIRQAAEAHPPVGRGTGKRVLTAKERKTQIDDRN
KLTEHFIITLPMLLSKYSADAEKVANLLQIPQYFDLEIYSTGRMEKHLDALLKQIKFVVEKHVESDV
LEACSKTYSILCSEEYTIQNRVDIARSQLIDEFVDRFNHSVEDLLQEGEEADDDDIYNVLSTLKRL
TSFHNAHDLTKWDLFGNCYRLLKTGIEHGAMPEQIVVQALQCSHYSILWQLVKITDGSPSKEDLL
VLRKTVKSFLAVCQQCLSNVNTPVKEQAFMLLCDLLMIFSHQLMTGGREGLQPLVFNPDTGLQS
ELLSFVMDHVFIDQDEENQSMEGDEEDEANKIEALHKRRNLLAAFSKLIIYDIVDMHAAADIFKHY
MKYYNDYGDIIKETLSKTRQIDKIQCAKTLILSLQQLFNELVQEQGPNLDRTSAHVSGIKELARRF
ALTFGLDQIKTREAVATLHKDGIEFAFKYQNQKGQEYPPPNLAFLEVLSEFSSKLLRQDKKTVHS
YLEKFLTEQMMERREDVWLPLISYRNSLVTGGEDDRMSVNSGSSSSKTSSVRNKKGRPPLHKK
RVEDESLDNTWLNRTDTMIQTPGPLPAPQLTSTVLRENSRPMGDQIQEPESEHGSEPDFLHNP
QMQISWLGQPKLEDLNRKDRTGMNYMKVRTGVRHAVRGLMEEDAEPIFEDVMMSSRSQLEDM
NEEFEDTMVIDLPPSRNRRERAELRPDFFDSAAIIEDDSGFGMPMFGSGSGGMGDSLSWLLRLL
N

SA2

10xHis – SA2 WT – ybbR

MHHHHHHHHHHGSGSSGIAAPEIPTDFNLLQESETHFSSDTDFEDIEGKNQKQGKGKTCKKGK
KGPAEKGKGGNGGGKPPSGPNRMNGHHQQNGVENMMLFEVVKMGKSAMQSVVDDWIESYK
HDRDIALLDLINFFIQCSGCKGVVTAEMFRHMQNSEIIRKMTEEFDEDSGDYPLTMAGPQWKKF
KSSFCEFIGVLVRQCQYSIIYDEYMMDTVISLLTGLSDSQVRAFRHTSTLAAMKLMTALVNVALNL
SINMDNTQRQYEAERNKMIGKRANERLELLLQKRKELQENQDEIENMMNAIFKGVFVHRYRDAI
AEIRAICIEEIGIWMKMYSDAFLNDSYLKYVGWTMHDKQGEVRLKCLTALQGLYYNKELNSKLEL
FTSRFKDRIVSMTLDKEYDVAVQAIKLLTLVLQSSEEVLTAEDCENVYHLVYSAHRPVAVAAGEF
LYKKLFSRRDPEEDGMMKRRGRQGPNANLVKTLVFFFLESELHEHAAYLVDSMWDCATELLKD
WECMNSLLLEEPLSGEEALTDRQESALIEIMLCTIRQAAECHPPVGRGTGKRVLTAKEKKTQLD
DRTKITELFAVALPQLLAKYSVDAEKVTNLLQLPQYFDLEIYTTGRLEKHLDALLRQIRNIVEKHTD
TDVLEACSKTYHALCNEEFTIFNRVDISRSQLIDELADKFNRLLEDFLQEGEEPDEDDAYQVLSTL
KRITAFHNAHDLSKWDLFACNYKLLKTGIENGDMPEQIVIHALQCTHYVILWQLAKITESSSTKED
LLRLKKQMRVFCQICQHYLTNVNTTVKEQAFTILCDILMIFSHQIMSGGRDMLEPLVYTPDSSLQS
ELLSFILDHVFIEQDDDNNSADGQQEDEASKIEALHKRRNLLAAFCKLIVYTVVEMNTAADIFKQY
MKYYNDYGDIIKETMSKTRQIDKIQCAKTLILSLQQLFNEMIQENGYNFDRSSSTFSGIKELARRF
ALTFGLDQLKTREAIAMLHKDGIEFAFKEPNPQGESHPPLNLAFLDILSEFSSKLLRQDKRTVYVY

LEKFMTFQMSLRREDVWLPLMSYRNSLLAGGDDDTMSVISGISSRGSTVRSKKSKPSTGKRKV
VEGMQLSLTEESSSSDSMWLSREQTLHTPVMMQTPQLTSTIMREPKRLRPEDSFMSVYPMQT
EHHQTPLDYNRRGTSLMEDDEEPIVEDVMMSSEGRIEDLNEGMDFDTMDIDLPPSKNRRERTE
LKPDFFDPASIMDESVLGVSMFGSSGDSLEFIASKLA

Rad21
6x His – Rad21 (Δ1-280; Δ421-631) – TEV site – MBPx

MGSSHHHHHHSGVDPVEPMPTMTDQTTLVPNEEEAFALEPIDITVKETKAKRKRKLIVDSVKEL
DSKTIRAQLSDYSDIVTTLDLAPPTKKLMMWKETGGVEKLFSLPAQPLWNNRLLKLFTRCLTPLV
PEDLRKRRKGGEADNLDEFLKEFGSSGENLYFQGGSGSNSSSGGSGGGSGKIEEGKLVIWING
DKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLA
EITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKG
KSALMFNLQEPYFTWPLIAADGGYAFKYGDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSI
AEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELA
KEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRVAATMENAQKGEIMPNIPQMSAFW
YAVRTAVINAASGRQTVDEALKDAQT

*Appendix - switchSENSE® measurements of CTCF target specificity*

In this appendix, I will present experiments carried out at the switch sense *switchSENSE®* instrument from Dynamic Biosensors GmbH with support from Veronika Bobinger (Dynamic Biosensors GmbH) and Dr. Günther Woehlke (TUM). SwitchSENSE® measurements can provide accurate binding affinities and kinetics of DNA-binding proteins to different oligonucleotide substrates. With this approach, I verified the site-specific binding of recombinantly purified CTCF to the designed CTCF consensus motif, which set the basis for single-molecule experiments in the DNA curtains set-up **(Chapter II 4.)**.

**Material and methods**

To perform switchSENSE® measurements, DNA strands (see below) containing a sequence motif (core motif: blue, upstream motif: green) as well as a single-stranded DNA (ssDNA) overhang (underlined) were mixed to a concentration of 5 µM and heated up to 95 °C, before slowly cooling down to RT at 1 °C per min in TE50 buffer (50 mM NaCl, 10 mM Tris pH 7.5, 1 mM EDTA). Afterwards, 250 nM double-stranded DNA (dsDNA) was hybridised to 200 nM fluorescently labelled Ra1 or Ra2 DNA nanolever from Dynamic Biosensors GmbH (AS-1-Ra, AS-2-Ra) **(Figure 7A)** via the ssDNA overhang by incubation at 25 °C for 1 h in the dark.

**Control DNA:**

5'-GGT AGC GAG TGA GAC TAC TAC ATC GCT TCA GGA TCG ACA AGA TCT GTA CAC ACA GAC CGA GTA G-3'

5'- CTA CTC GGT CTG TGT GTA CAG ATC TTG TCG ATC CTG AAG CGA TGT AGT AGT CTC ACT CGC TAC CAT CAG TAC TTG TCA ACA CGA GCA GCC CGT ATA TTC TCC TAC AGC ACT A -3'

**CTCF site:**

5'-CGA TCG GTG ATG ACG ATA GGC TTT TTG GTG CCC TCT GCT GGC CAG TTT TGG AAC TGC AGT TTA G-3'

5'-CTA AAC TGC AGT TCC AAA ACT GGC CAG CAG AGG GCA CCA AAA AGC CTA TCG TCA TCA CCG ATC GAT CAG TAC TTG TCA ACA CGA GCA GCC CGT ATA TTC TCC TAC AGC ACT A -3'

**Core motif:**

5'-CGA TCG GTG ATG ACG ATA GGC TTT TTG GTG CCC TCT GCT GGC CCT GTA CAC ACA GAC CGT TTA G-3'

5'-CTA AAC GGT CTG TGT GTA CAG GGC CAG CAG AGG GCA CCA AAA AGC CTA TCG TCA TCA CCG ATC GAT CAG TAC TTG TCA ACA CGA GCA GCC CGT ATA TTC TCC TAC AGC ACT A-3'

**Upstream motif:**

5'-CGA TCG GTG ATG ACG ATA GGC TTC GCT TCA GGA TCG ACA AGA TAG TTT TGG AAC TGC AGT TTA G-3'

5'-CTA AAC TGC AGT TCC AAA ACT ATC TTG TCG ATC CTG AAG CGA AGC CTA TCG TCA TCA CCG ATC GAT CAG TAC TTG TCA ACA CGA GCA GCC CGT ATA TTC TCC TAC AGC ACT A-3'

One Ra1- and one Ra2-annealed oligo were mixed to a final DNA concentration of 100 nM each. The switchSENSE® gold chip (ADP-48-2-0) contains two measurement chambers with complementary ssDNAs 1/2 for Ra1 or Ra2, respectively, enabling two simultaneous

measurements. Before each switchSENSE® experiment, a Chip-test was carried out to ensure the proper functioning of the gold surface. The gold surface was then functionalised by hybridisation of Ra1 and Ra2 DNA to short ssDNAs. Measurement chambers were treated with passivation solution (Dynamic Biosensors GmbH SOL-PAS-1-5) and then equilibrated with CTCF measurement buffer (25 mM Hepes pH 7.8, 225/300 mM NaCl, 5 % glycerol, 0.05 % Tween-20, 100 µM $ZnCl_2$, 1 mM $MgCl_2$, 1 mg/ml BSA). For 2 minutes, 100 nM CTCF wild-type (WT) or ΔRBR was incubated in CTCF measurement buffer in the measurement chamber. Next, CTCF dissociation was monitored for 10 minutes by incubation using CTCF measurement buffer without protein. Due to CTCF surface sticking, the measurement chambers were treated with 6 M guanidinium chloride after each measurement and a new round of measurements was started with a new functionalisation with Ra1 or Ra2 DNA nanolevers.

Binding and unbinding curves were fitted with exponential curves to determine $k_{on}$ (app) and $k_{off}$ and finally $k_{on}$ **(equation (2))** with protein concentration c(prot). WT unbinding from the sequence motifs was fitted with a double exponential fit to correct for a buffer jump. $K_D$ was determined from $k_{off}$ and $k_{on}$ **(equation (3))**.

$$k_{on} = \frac{k_{on}(app) - k_{off}}{c(prot)} \quad (2)$$

$$K_D = \frac{k_{off}}{k_{on}} \quad (3)$$

**Results and discussion**

In order to test both the functionality of my CBS designed from the CTCF consensus motif (204) and of my purified and fluorescently labelled CTCF WT and CTCF ΔRBR constructs, I performed affinity measurements on a helix *switchSENSE*® setup from Dynamic Biosensors GmbH. Thereby the respective CTCF construct is flushed into a flow chamber **(Figure 7A,B)** - previously functionalised with fluorescently labelled DNA - and binding kinetics can be observed due to quenching of the fluorophore by CTCF association. After an equilibrium is reached, unbinding is recorded by flushing in the measurement buffer without protein. Due to problems with CTCF sticking to the gold surface at low salt concentrations, relatively high salt concentrations were chosen for WT (300 mM NaCl) and ΔRBR (225 mM NaCl) **(Figure 8A)**. Additionally, the surface was freshly functionalised after each measurement, since surface sticking led to a remaining quenching signal even at the end of the measurement **(Figure 8A)**.

Affinities **(Figure 8B)** for CTCF WT (top) and ΔRBR (bottom) for the complete CBS and the core motif were calculated from binding and unbinding kinetics according to formulas (2) and (3). The measured $K_D$-values (249.5 ± 166 nM WT at 300 mM NaCl, 191.0 ± 16 nM ΔRBR at 225 mM NaCl) are thereby much higher than concentrations of CTCF used for single-molecule measurements (1-20 nM). However, for single-molecule measurements, CTCF was bound at 50 mM NaCl and only afterwards, salt concentration was increased to 300 mM to enrich CTCF on its target sites, the lower salt concentration explaining why CTCF was able to bind at lower protein concentrations.

Verifying the CTCF consensus sequence

Binding to the designed CBS or the CTCF core motif **(Figure 7A)** was observed for WT and ΔRBR under given salt concentrations **(Figure 8A)**. Surprisingly, the affinity for the complete CBS compared to the core motif was slightly higher for ΔRBR ($p = 0.0092$, z-test) but not for the WT ($p = 0.83$, z-test), even though the WT contains more of the outer ZFs interacting with the upstream motif (30). No binding was observed for DNA containing only the CTCF upstream motif or to control DNA. This tells us, that my designed sequence from the CTCF consensus motif (204) is functional for site-specific CTCF binding and was cloned into λ-DNA for DNA curtains measurements.

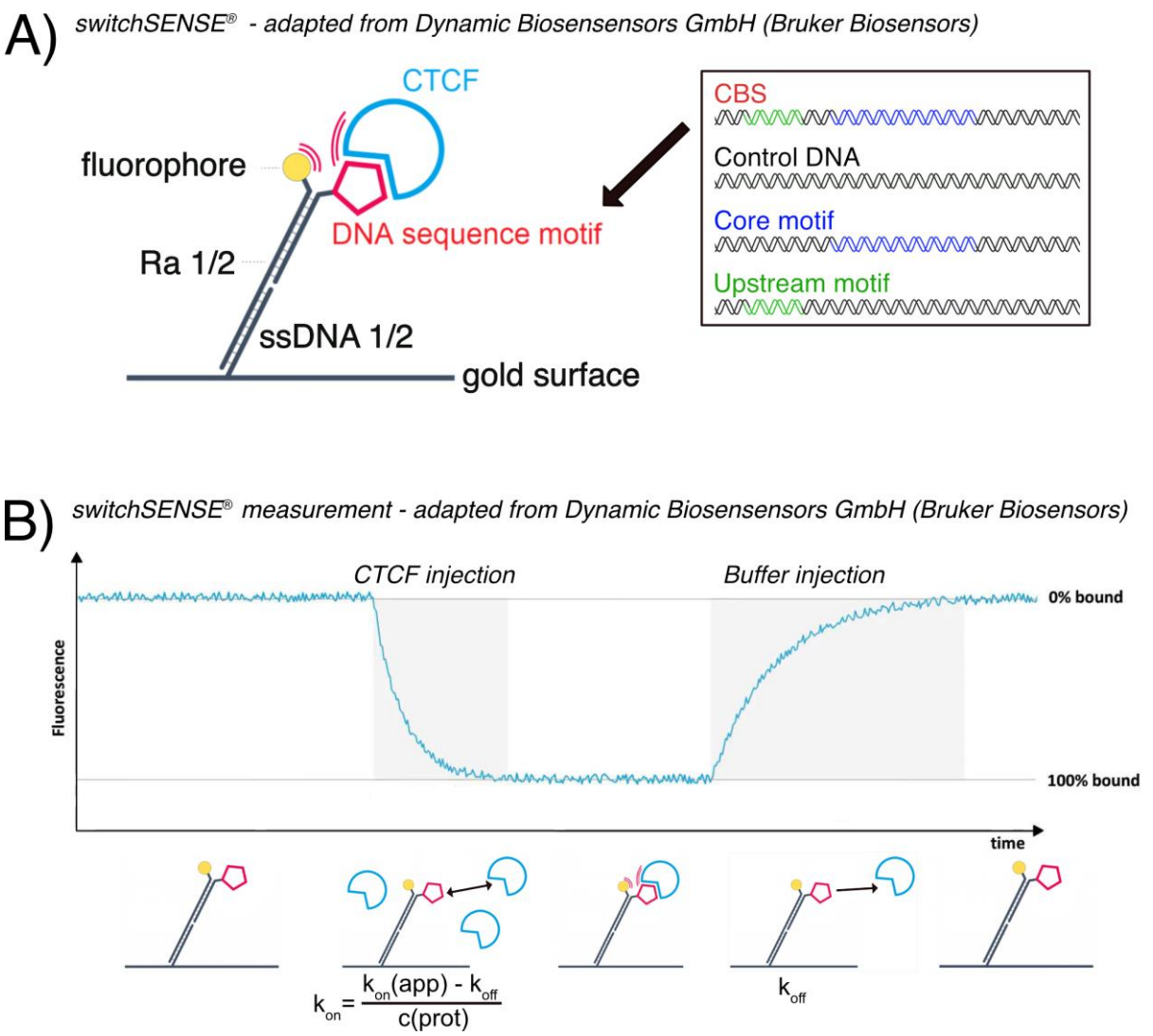



*Figure 7: Protein-DNA binding kinetics measured with switchSENSE® (Dynamic Biosensors GmbH). A) Gold chips contain two measurement chambers with ssDNA 1 or 2 attached. Two DNA sequence motifs previously annealed to DNA nanolever Ra1 or Ra2 can be measured simultaneously. Both Ra1 and Ra2 contain an attached fluorophore. DNA fluorescence is quenched upon protein binding to the DNA sequence motif. Four different motifs containing the CBS, control DNA, the core motif or the upstream motif were measured. Figure adapted from Dynamic Biosensensors GmbH (Bruker Biosensors) (29.09.2024) (https://www.dynamic-biosensors.com/switchsense/) (581). B) During the measurement, CTCF is injected, leading to a decrease in fluorescence and association kinetics are monitored until an equilibrium is reached. Afterwards, CTCF measurement buffer without protein is flushed in, leading to an increase in fluorescence due to CTCF unbinding. Dissociation kinetics are monitored until the protein is unbound. Figure adapted from Dynamic Biosensensors GmbH (Bruker Biosensors) (29.09.2024) (https://www.youtube.com/watch?v=RxIonG2cRJU switchSENSE® Dynamic Response Sensing 0:52 Bruker Biosensors) (582).*

Central and outer ZFs are important for binding specificity

According to the *switchSENSE®* measurements, the effect of the upstream motif seems to be negligible for binding specificity. This agrees well with previous data showing CTCF to have a higher affinity for the core motif compared to the upstream motif (583). In line, the ZF9-CT construct containing only the upstream motif-binding ZFs was not enriched on the CBS during my single-molecule measurements **(Chapter II 4., Figure 2)**, but instead bound to AT-rich regions. However,

the ZF4-7 construct containing only the core motif-binding ZFs was not enriched either **(Chapter II 4., Figure 2)**. The ΔRBR mutant missing only two of the three upstream motif-binding ZFs and one downstream DNA-binding ZF (30) and the CTCF WT with all eleven ZFs show CBS-binding specificity in bulk as well as in my single-molecule data **(Chapter II 4., Figure 1,2)**, showing that a combination of core and outer motif-binding ZFs is required for CBS recognition.

Additionally, the ΔRBR mutant displayed a lower lifetime than the WT in single-molecule assays **(Chapter II 4., Figure 2)**. The outer ZFs, therefore, seem to be required to additionally stabilise the core motif-binding ZFs. Although CTCF's ZF9-ZF11 bind to a distinct upstream motif *in vivo* (204), I observed instead binding to AT-rich regions for the ZF9-CT construct **(Chapter II 4., Figure 2)**, suggesting that the narrower minor grooves of AT-rich DNA (76, 544) might increase binding affinity for arginine and lysine residues in CTCF's outer ZFs. The fact that *in vivo,* some binding sites contain an additional upstream motif while others do not (204) might be due to CTCF's interactions with additional binding partners interacting with flanking DNA like, for example, the SA subunit of the cohesin complex **(Chapter II 4. Figure 3)**. I conclude that the core motif-binding ZFs are stabilised by additional outer ZFs without the requirement for an additional sequence motif.

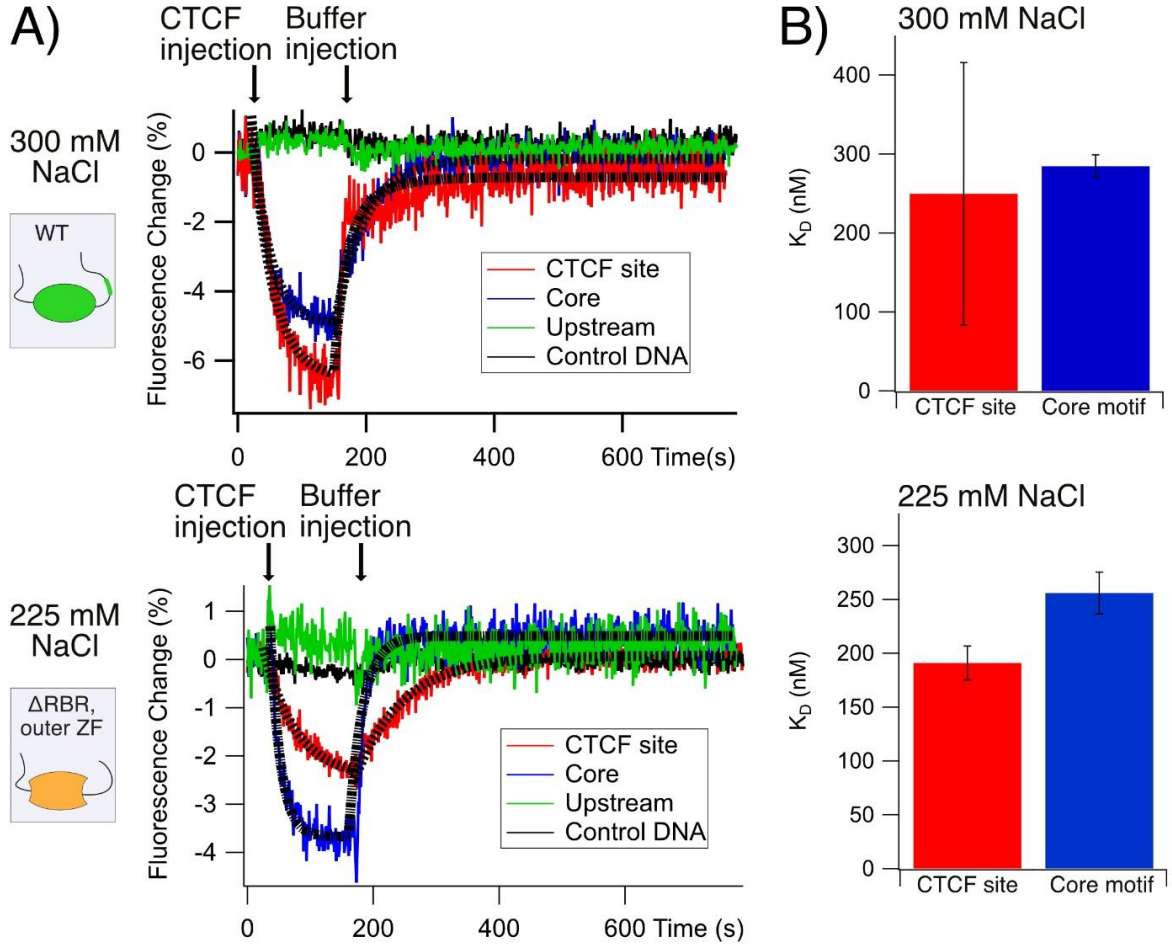



*Figure 8: CTCF WT and ΔRBR bind the core motif and CBS but not the upstream motif. A) Fluorescence intensity over time for measurements with CTCF WT (top) and ΔRBR (bottom) on different DNA sequence motifs (CBS, control DNA, core motif and upstream motif in red, black, blue and green respectively). Binding and unbinding were fitted with a single exponential. In the case of the CTCF WT – CTCF site a bimodal unbinding was observed due to a buffer jump and fitted with a double exponential. Fits are displayed as black dashed lines. B) KD of CTCF WT (top) and ΔRBR (bottom) on a complete CTCF site (red) or the core motif (blue) at 300 mM NaCl or 225 mM NaCl respectively.*

## II 5. CTCF forms oligomeric structures with and without nucleotides

**Summary:**
The transcriptional regulator CTCF controls interaction frequency between regulatory DNA domains enhancing interactions by bridging DNA segments, which are far apart on the DNA strand, while also being able to prevent interactions of close-by domains by acting as a boundary factor. Both of these processes have been linked to CTCF oligomerisation. This oligomerisation might enable CTCF to form phase-separated condensates since it contains both intrinsically disordered regions as well as nucleic acid-binding domains commonly associated with droplet formation. CTCF's role in nuclear phase separation depends on the interaction with additional proteins and nucleic acids. Here I analyse if CTCF can form oligomers and phase-separated droplets on its own *in vitro* and how this influences its DNA-binding properties.

Mass photometry experiments display that CTCF forms oligomeric structures on its own in the absence of both DNA and RNA. Furthermore, oligomerisation occurred at physiological salt and protein concentrations. Truncation mutants of CTCF revealed that this oligomerisation depends on CTCF's zinc fingers (ZFs) but not on the unstructured termini.

Facilitated by the molecular crowder polyethylene glycol (PEG)-8000 CTCF was also able to form mobile droplets at physiological conditions. However, the droplets were both salt- and 1,6-hexanediol stable and did not fuse. Therefore, I assume that CTCF cannot perform phase separation, but CTCF oligomers might instead act as an interaction hub for other proteins involved in transcriptional regulation and genome architecture.

I tested if CTCF oligomers interact with nucleic acids. AFM imaging revealed that CTCF oligomers integrate DNA into their structures. Furthermore, when I overexpressed CTCF in eukaryotic cells it formed clusters containing nucleic acids. CTCF's ability to form oligomers might therefore enable it to form larger complexes with DNA and thereby spatially control nuclear interactions.

**Author contributions:**
Jonas Huber performed protein purifications and bright field microscopy. Sarah Zernia performed mass photometry measurements. Michael Scheckenbach and Johann Bohlen performed AFM measurements. Jonas Huber wrote the manuscript with help from all other authors.

**Abstract:**
The process of phase separation is fundamental for all organisms, regulating cellular reactions by separating specific proteins and other biomolecules from the surrounding liquid. Inside the nucleus, phase separation modifies chromatin density, thereby controlling DNA accessibility for regulatory factors involved for example in DNA repair and gene expression. The transcriptional insulator CTCF is assumed to participate in the formation of nuclear condensates, especially in the context of transcription by RNA Pol II. CTCF contains eleven nucleic acid-binding domains as well as two unstructured termini, both features that are known to promote phase separation. Here mass photometry was applied to show that CTCF can form oligomers at physiological protein- and salt concentrations independent of nucleic acids or other proteins. Analysis of different CTCF constructs thereby revealed that the concentration-dependent oligomerisation relies on the ZFs but not the unstructured termini. These oligomers formed droplets under crowded conditions and were able to form clusters with DNA in AFM measurements. CTCF also formed clusters with nucleic acids when overexpressed in eukaryotic cells. CTCF oligomerisation might, therefore, enable it to form compact DNA structures inside the nucleus and regulate genomic contacts in a cohesin-independent manner.

## *5.1 Introduction*

Compartmentalisation plays an important role in many biological processes, separating biomolecules and thereby only allowing specific interactions and reactions to occur at certain time points (584). The most prominent example is the cell, which is further compartmentalised into membrane-bound or membrane-less structures called organelles (26, 584). As an example of membrane-bound organelles, the nucleus separates transcription from translation during protein biosynthesis and allows for the regulation and modification of RNA transcripts by RBPs (26, 585). Multiple membrane-less compartments exist within the nucleus (e.g. nucleolus, paraspeckles, nuclear stress bodies, Cajal bodies, cleavage bodies, histone locus bodies…), which are important in controlling transcription, regulating mRNA splicing and modifying protein functions and properties by posttranslational modifications (PTMs) (26, 28, 340, 348).

The physiochemical forces that drive the formation of membrane-less organelles have remained unknown for a long time until in 2009 it was shown that nucleoli are liquid phase-separated droplets formed by RNA and proteins (26, 586). Phase separation thereby mostly depends on intrinsically disordered regions (IDRs) and prion-like domains involved in protein homo-oligomerisation and on the formation of larger networks with nucleic acids (315, 345, 587–589). For phase separation to occur the energetic gain from these interactions must be larger than the loss of entropy from forming segregated phases, which depends on the concentration of these macromolecules as well as on the physical and chemical properties of the surrounding liquid (28, 317). Phase separation allows the formation of separate reaction rooms inside the nucleus, for example of highly and weakly expressed genomic regions ensuring efficient gene expression in euchromatin, while at the same time upkeeping genome stability by preventing expression in for example centromeric or transposon-rich heterochromatin regions (23, 102, 312, 313, 590).

Different kinds of phase separation have been defined based on the physiochemical properties of the separated phase (318). For liquid-liquid phase separation (LLPS), multiple electrostatic interactions between proteins and nucleic acids lead to the formation of a droplet with the phase properties of a liquid (318). This enables the molecules to quickly diffuse and interact with each other within the droplet while also leading to reduced interactions with molecules outside of the droplet (318). Another example of a phase-separation process is liquid-gel phase separation (LGPS), which involves protein or polysaccharide chains crosslinked into networks that incorporate

large amounts of water and form gel-like structures used for different applications in cell cultures, biomedicine and food industry (318, 591–593). A third type of phase separation is polymer-polymer phase separation (PPPS), which appears when attractive forces between polymer chain segments, for example between multiple DNA-bound proteins in the formation of dense heterochromatin, are stronger than interactions with the surrounding liquid, causing a strong compaction of the polymer segments (318, 594). LGPS and PPPS display slower diffusion within the separated phase than LLPS but faster exchange with the environment (318). However, droplets formed by either of the three can merge to form larger droplets (318).

One prominent example of nuclear condensates are Pol II and transcriptional regulator containing super-enhancers, controlling the expression of a large number of genes (313, 351). These condensates are formed by the IDRs of the transcriptional coactivator Mediator as well as by Pol II (25, 347, 350), which can form condensates via its intrinsically disordered C-terminal domain (25, 340) and other IDR-containing coactivators like BRD4 (319). Genes controlled by super-enhancers are often positioned between CTCF-binding sites (CBSs) (354). Deletion of these CBSs leads to increased expression of non-target genes outside of CTCF-mediated super-enhancer boundaries, which implies that CTCF is important for regulating super-enhancer activity (173, 174, 595). However, CTCF sometimes forms imperfect boundaries with leaky expression even in the presence of CBSs (173). Whether this regulation is mediated by monomeric or oligomeric CTCF and if its caused by CTCF's acting as a boundary factor on its CBS or by CTCF directly influencing the condensate formation properties of these super-enhancers remain interesting open questions.

CTCF contains eleven nucleic acid-binding domains, disordered regions and different PTMs (29, 30, 225, 248, 250), which are all structures commonly found among droplet forming RNA-binding proteins (RBPs) (28), suggesting a role of CTCF in nuclear phase separation. Furthermore, CTCF droplet formation has already been observed *in vitro* (221, 224, 430) as well as *in vivo* (221, 224). However, how CTCF-nucleic acid interactions, CTCF oligomerisation and its interaction with other proteins influence droplet properties is still under debate.

This study shows that CTCF forms oligomers in a concentration-dependent manner, even in the absence of nucleic acids or other cofactors. These oligomers form droplets under crowded conditions at physiological salt and protein concentrations. AFM reveals that CTCF oligomers form large clusters with DNA. Additionally, CTCF formed clusters with nucleic acids when overexpressed in eukaryotic cells. By creating truncation-mutants of CTCF's unstructured regions or ZFs, it is shown that CTCF oligomerisation mainly depends on its ZFs.

## *5.2 Material and methods*

**CTCF-cluster purification**

Monomeric CTCF wild-type (WT) and CTCF mutant purifications and labelling was performed according to a previous publication (229). Here an alternative purification strategy was also applied to obtain CTCF WT clusters instead of monomeric CTCF WT. In this case, a CTCF construct containing an N-terminal glutathione S-transferase (GST)-tag followed by an HRV3C cleavage site and a Flag-tag was used. Cell pellets were resuspended in cluster resuspension buffer (25 mM Hepes pH 7.9, 200 mM NaCl, 150 mM KCl, 5 % glycerol, 0.05 % Tween, 100 µM $ZnCl_2$, 1 mM TCEP, 1 mM PMSF, protease inhibitor tablet (Roche), 10 µg/ml RNase (Roth 7156.1)). Cells were sonicated for 60 sec at 20 % amplitude and 40 % duty cycle. In contrast to purification for monomeric CTCF, here, no room temperature pierce nuclease digestion step at low salt allowing for efficient nucleic acid cleavage was applied after sonication. After centrifugation (17 krpm, 4 °C, 30 min) followed by ultracentrifugation (42 krpm, 4 °C, 1 h), the supernatant was filtered through 0.45 µm and 0.22 µm polyethersulfone (PES) filters (Fisher Scientific). The supernatant was then

applied to a Cytiva 5 ml GS-trap previously equilibrated in cluster resuspension buffer. After a 25 ml cluster resuspension buffer wash step, the protein was eluted in a 25 CV gradient to 20 mM glutathione without a high salt wash step. CTCF was dialysed for 1 h against HRV3C cleavage buffer (50 mM Tris pH 7.0, 150 mM NaCl, 5 % Glycerol, 0.05 % Tween, 1 mM TCEP) containing 100 µM ZnCl2, and 1h HRV3C cleavage buffer without $ZnCl_2$ before concentration to 100 µl using 30 kilodalton (kDa) MWCO amicons and cleavage overnight at 4 °C using 100 µg HRV3C protease (made in house). The protein sample was further purified using a superose 6 size-exclusion column (Cytiva) in cluster resuspension buffer without protease inhibitors. Samples were stored at -70 °C.

**Mass photometry**

Mass photometry measurements were performed as described previously (229). A histogram with 60 bins was calculated from all measured masses and Gaussian fits were applied to individual protein peaks using Igor Pro 8 (Wavemetrics).

**Bright-field microscopy**

For the preparation of a measurement chamber, borosilicate glass slides were cleaned by rinsing with millipore water and dried with $N_2$. Two 1 cm wide parafilm stripes were arranged on top of the glass forming a 1 cm wide channel. Parafilm was melted on a hot plate for 20 sec at 130 °C and a coverslip was fixed on top to cover the chamber. CTCF was diluted in phase separation buffer (25 mM Hepes pH 8.3, 0.05 % Tween, 5 % Glycerol, 100 µM $ZnCl_2$, 150 mM NaCl, 1 mM DTT) before 350 mg/ml PEG solution was added to final concentrations of 90 – 3000 nM CTCF and 10 or 100 mg/ml PEG. After 5 min incubation time droplet stability was tested by increasing the salt concentration from 150 to 2 M NaCl or addition of 10 % 1,6-hexanediol (Sigma-Aldrich). Droplet formation was analysed in 20 s videos using a Nikon Eclipse Ti2 with a small white light placed above the microscope, the electron multiplier gain set to 0 and the illumination time to 1 s. In some cases, CTCF labelled with Halo-Tag TMR (Promega #G8251) was used. In these cases, videos were recorded as described above but instead of using a white light a Coherent® OBIS™ 1280720 | 561 nm LS 150 mW laser was used at 10 mW (0.1 µW/µm$^2$).

**Atomic-force microscopy**

AFM measurements were performed similarly to a previous protocol (596) on a NanoWizard® 3 ultra AFM (JPK Instruments AG). The mica surface (Quality V1, Plano GmbH) was incubated with a 10 mM solution of $NiCl_2$ for 3 minutes followed by three ultrapure water washing steps. After drying with $N_2$, the surface was then incubated for 3 min with 10 nM unlabelled CTCF WT, 2.5 nM 944 bp DNA containing four CBSs (two times 5'-TGCAGTTCCAAAACTGGCCAGCAGAGGGC ACCAAA-3' and two times 5'-TGCAGTTCCAAAAGCGGCCAGCAGGGGGCGCCCAA-3' spaced by 129 bp each), or a mixture of CTCF WT and DNA in imaging buffer (40 mM Tris pH 7.5, 20 mM NaCl, 12.5 mM $MgCl_2$). This was followed by three washing steps with 300 µl imaging buffer before measurements were performed in 1.5 ml imaging buffer. A BioLeverMini cantilever ($v_{res}$ = 110 kHz air / 25 kHz fluid, $k_{spring}$ = 0.1 N/m, Bruker AFM Probes) was used to scan different area sizes between 0.4 x 0.4 µm and 3 x 3 µm in AC mode. Cross-section heights of characteristic structures were analysed using the software JPKSPM data processing.

## *5.3 Results*

**CTCF oligomerisation is concentration-dependent**

CTCF clustering has been observed *in vitro* (221, 222, 225) and *in vivo* (221, 223, 353) and is linked to genome organisation (223, 226), transcription (223), DNA compaction (221) and condensate formation (224, 353). To test at which conditions CTCF forms oligomers we performed mass photometry measurements of CTCF WT and four different CTCF variants at two different protein concentrations (2 nM and 20 nM) **(Figure 1).**

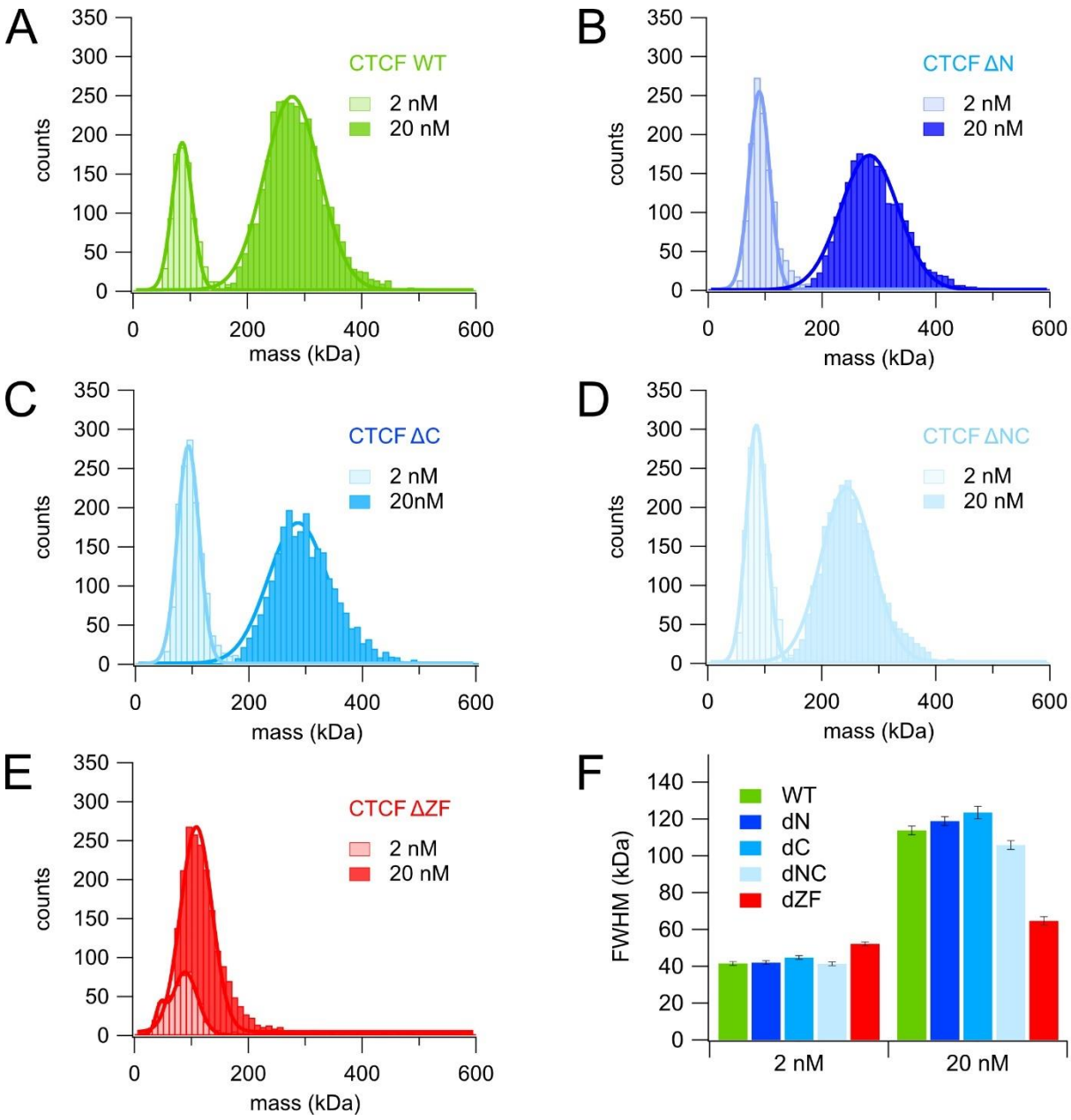



*Figure 1: CTCF oligomerisation depends on protein concentration and CTCF's ZFs. A)-E) Mass photometry of CTCF variants. Histograms were fitted with a multi-Gauss function. F) FWHM (full-width at half maximum) of Gauss fits of mass photometry data of CTCF variants.*

At 2 nM concentration, CTCF WT displays a normal distribution with a mean of 85.9 ± 0.4 kDa **(Figure 1A)** and an FWHM of 41.5 ± 0.4 kDa **(Figure 1A,F),** which can be assigned to monomeric CTCF (122.2 kDa). In contrast at 20 nM concentration, CTCF WT displays a much broader single peak (FWHM = 113.7 ± 2.4 kDa) at a much higher molecular weight of 278.2 kDa ± 0.9 kDa. The shift to a mass much higher than monomeric CTCF therefore suggests oligomerisation and the peak width the presence of multiple oligomeric states, which cannot be distinguished from each other. CTCF WT is therefore able to form oligomers at a concentration of at least 20 nM, which is around 8 times lower than CTCF's concentration inside the nucleus (597). To determine which domains are required for CTCF oligomerisation, measurements were repeated with different truncation mutants. Since IDRs are often involved in phase separation processes (26), the role of CTCF's unstructured termini was analysed. Surprisingly, mutants missing one or both termini **(Figure 1B-D)** displayed very similar behaviour compared to the CTCF WT. At 2 nM concentration, the variants displayed a mass peak at monomeric protein size (CTCF ΔN: 89.9 ± 0.5 kDa/90.9 kDa,

CTCF ΔC: 93.2 ± 0.4 kDa/105.0 kDa, CTCF ΔNC: 85.2 ± 0.4 kDa/74.7 kDa measured/actual mass), which, at 20 nM, shifted to higher masses (CTCF ΔN: 283.1 ± 1.1 kDa, CTCF ΔC: 286.3 ± 1.3 kDa, CTCF ΔNC: 243.9 ± 0.9 kDa) **(Figure 1B-D)** as seen for the CTCF WT. This implies that these three mutants also form oligomers at sub-cellular concentrations and that the termini are not required for this process.

Instead, oligomerisation might happen via the CTCF's ZFs. Mass photometry measurements were hence repeated with a variant missing all eleven ZFs (ΔZF) **(Figure 1E)**. Indeed, only monomeric protein (84.2 kDa) at 2 nM (88.7 ± 0.4 kDa) and 20 nM (108.8 ± 0.9 kDa) was observed, and FWHM increased only by a minor extent (24 %), in contrast to all other constructs (for which FWHM increased by 156 – 183 %) **(Figure 1F).** Although the slight increase in molecular mass and FWHM might suggest that this mutant can also form oligomers, oligomerisation is clearly impeded compared to CTCF WT. Therefore, CTCF's ZFs play a vital role in CTCF oligomerisation.

In summary, CTCF WT can form oligomers at nuclear salt and protein concentrations without the requirement of additional cofactors like other proteins or nucleic acids and this oligomerisation mainly depends on its ZFs.

**CTCF forms mobile droplets under crowded conditions**

Having shown that CTCF forms oligomers at below physiological concentrations, I next analysed whether this oligomerisation leads to the formation of phase-separated droplets using bright-field microscopy.

To simulate the crowded environment in living cells, I added PEG 8000 as a molecular crowder to my CTCF samples. I observed the formation of mobile droplets with 3 µM CTCF **(Figure 2A)**. I used fluorescently labelled CTCF, to confirm that these droplets contain CTCF. Note that only the droplets which are in the focal plane display fluorescence.

Droplet formation was tested at different PEG- and CTCF concentrations **(Figure 2B)**. At low protein and PEG concentrations, no droplet formation was observed **(Figure 2B, left image)**. At 180 nM CTCF, which corresponds to the physiological CTCF concentration inside the nucleus during G1 (188 nM) and G2 (167 nM) (597), droplet formation was observed at a PEG-concentration of 100 mg/ml **(Figure 2B centre image).** At CTCF concentrations higher than 180 nM (e.g. 375 mM, 750 nM and 3 µM), droplets formed at both PEG concentrations. Higher protein and PEG concentrations led to increased droplet formation **(Figure 2B right image).**

To analyse the influence of electrostatic and hydrophobic interactions on droplet formation, 2 M NaCl or 10 % 1,6-hexanediol was added to pre-formed droplets. Interestingly, CTCF droplets remained stable after the addition of either NaCl or 1,6-hexanediol **(Figure 2C)**. This is surprising, as these additives are known to disrupt electrostatic or hydrophobic interactions, which made us question if the observed droplets actually resemble LLPS. In line with this, droplets were also unable to fuse in optical tweezer experiments (data not shown), which is another essential feature of all kinds of phase separation (318). The observed droplets might, therefore, not be condensates but large CTCF oligomers, scattering light in bright-field microscopy.

CTCF can, therefore, form large oligomers, which form droplets under crowded conditions at cellular salt and protein concentrations.

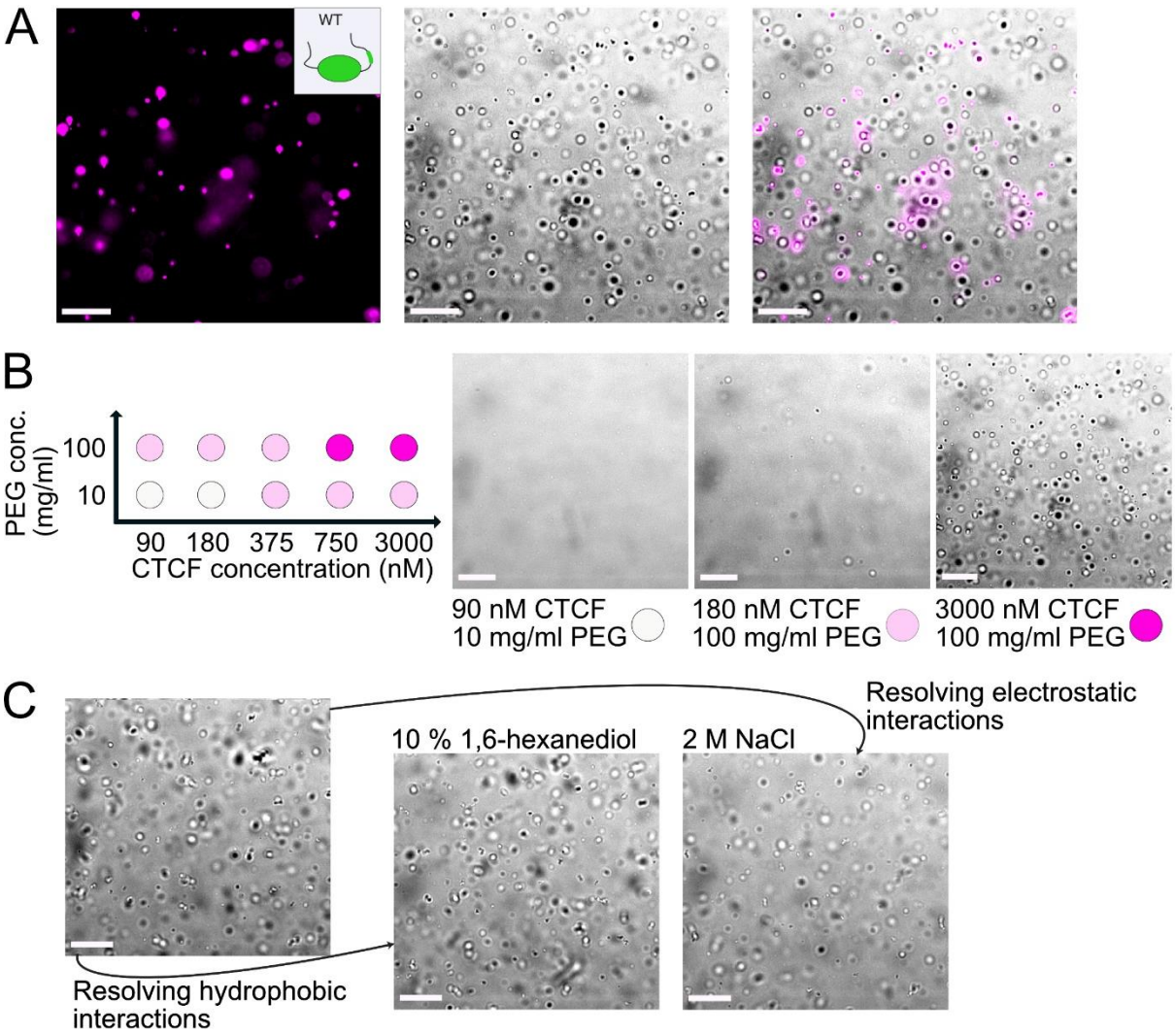



*Figure 2: CTCF forms droplets under physiological salt and protein concentrations. A) CTCF WT droplet formation was measured using a 561 nm laser (left) and bright-field microscopy (centre) and superimposed (right). The scheme of the CTCF WT is displayed in green. B) Phase diagram of CTCF WT droplet formation is displayed using transparent circles in case of no droplet formation (left picture), faint pink circles in case of weak droplet formation (central picture) and pink circles in case of strong droplet formation (right picture). C) Droplets were tested for stability by adding 10 % 1,6-hexanediol (left) or 2 M NaCl (right) after droplet formation. Droplets remain stable under both conditions. The size of the scale bar is 20 µm for all images.*

**CTCF integrates nucleic acids into oligomeric structures**

So far, it has been shown that CTCF can form oligomers and mobile droplets even in the absence of nucleic acids at physiological salt and protein concentrations. However, CTCF clustering has been linked to DNA compaction (221) and network formation with DNA (222). Therefore, the influence of nucleic acids on CTCF cluster formation was analysed. First, I investigated whether CTCF interacts with nucleic acids during protein expression in insect cells. Hence, I applied two different purification protocols to CTCF, with only one including a pierce universal nuclease treatment at RT, a high salt wash step and a heparin purification step, the other just the addition of RNAse A. In both cases, CTCF was afterwards further purified by size exclusion chromatography, and purity was analysed by SDS-PAGE as well as agarose gel electrophoresis.

For nuclease- and salt wash-treated CTCF WT, the chromatogram showed a single protein peak at 13.8 ml with an OD 260/280 ratio of 0.59, which suggests that CTCF could be purified as a monomer without significant nucleic acid contamination (598). **(Figure 3A)**. SDS-PAGE displayed

a single band between 150 kDa and 200 kDa, which is higher than the expected mass of 122.2 kDa. As CTCF-termini have been shown to behave like higher molecular weight structures in gel electrophoresis and size exclusion chromatography before (29), it is likely that full-length CTCF also runs differently. Alternatively, the higher molecular weight could be a cause of poly(ADP-ribosyl)ation (249). Purified CTCF was additionally analysed on a 1 % agarose gel stained with SYBR gold, which revealed no significant nucleic acid contamination.

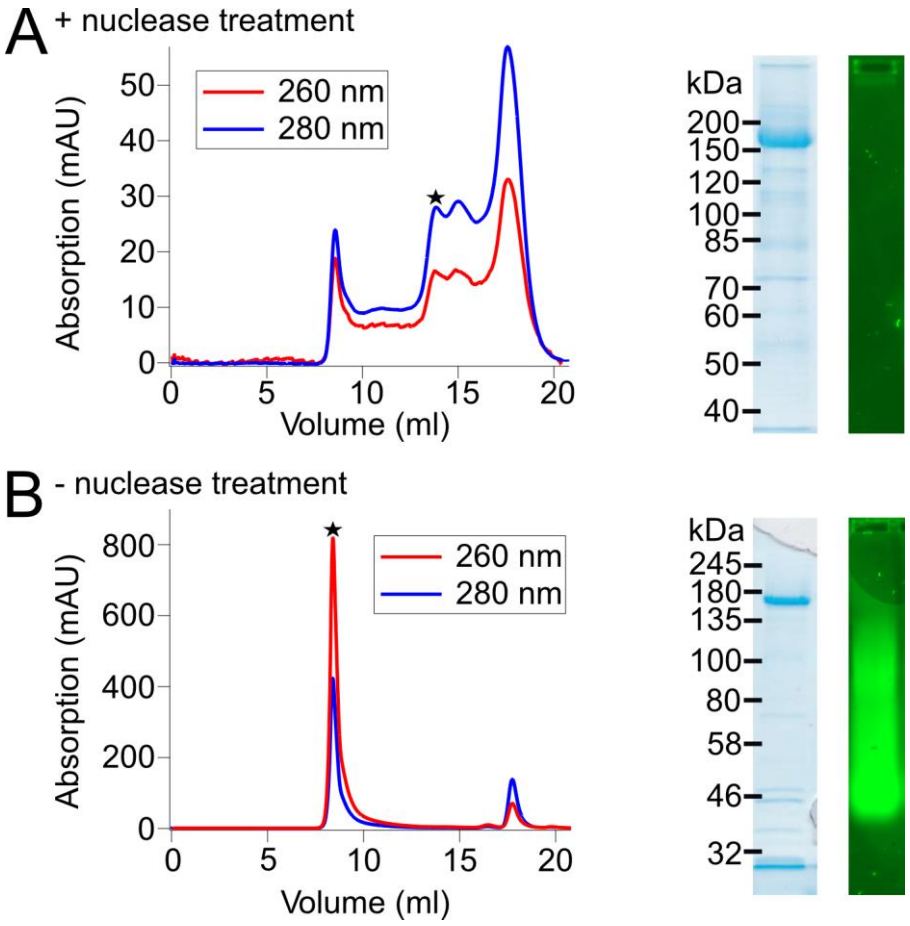



*Figure 3: CTCF binds nucleic acids in cell extracts. A) CTCF was purified with a pierce universal nuclease digestion step, nickel-column high salt wash, and heparin column purification before performing size-exclusion chromatography (left). Absorption at 260 nm (red) and 280 nm (blue) is plotted against the elution volume. SDS-gel (molecular weights of marker bands displayed by black bars) showing the CTCF peak fraction (labelled with a black star) is shown to the right. Peak fractions were analysed on 1 % agarose gel showing SYBR-gold staining in green. B) Same as A) but CTCF purification was performed without nuclease digestion, high salt wash and heparin purification step. CTCF elutes in the void peak during size-exclusion chromatography.*

In contrast, CTCF purified without salt wash and pierce universal nuclease treatment showed a drastically different size exclusion elution profile **(Figure 3B)**. No peak at 13.8 ml was observed. Instead, CTCF eluted in the void fraction, which was displayed by SDS-PAGE. The OD 260/280 ratio was 1.93, suggesting high nucleic acid contamination, which was confirmed by agarose gel electrophoresis and SYBR-gold staining. The presence of nucleic acids, therefore, led to cluster formation of CTCF, which resulted in an early elution from the gel filtration column. This shows that CTCF interacts and forms large clusters with nucleic acids in eukaryotic cells. I, therefore, wanted to test if nuclease-treated CTCF can form clusters with DNA *in vitro*.

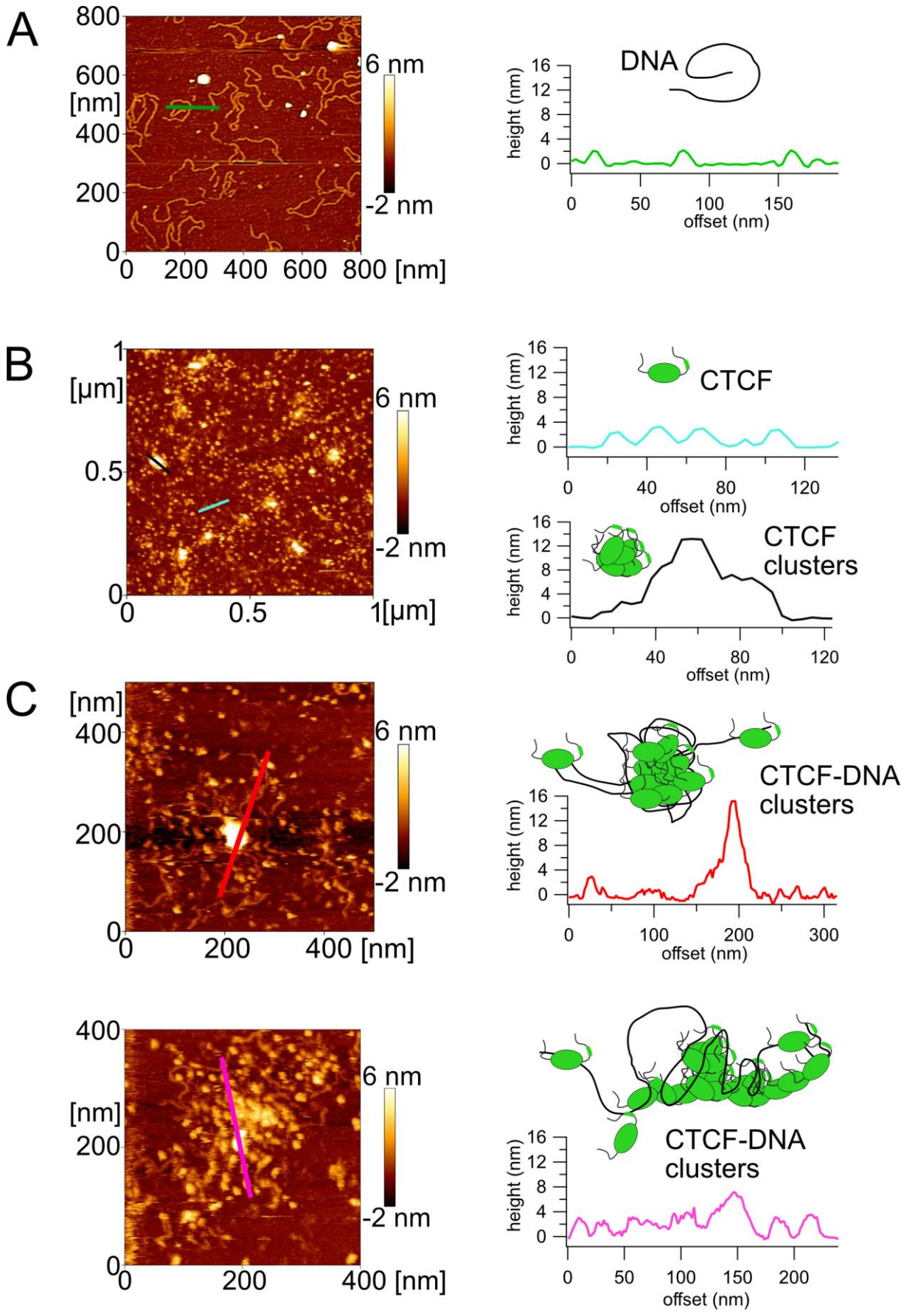



*Figure 4: CTCF oligomers form larger clusters with DNA. A) Representative AFM image for DNA alone. Positions at which a cross-section was taken are marked by a green line in the image, and the cross-section height is plotted to the right of the image. B) Same as A) but for CTCF alone.*

*Cross-sections of CTCF monomers (cyan line) and oligomers (black line) were taken and cross-section heights were plotted to the right. C) same as A) but for DNA and CTCF combined. CTCF forms clusters with DNA as shown by cross-sections displayed in the images by red and pink lines with cross-section height plotted to the right.*

AFM measurements were performed to further characterise CTCF oligomerisation. Thereby either DNA alone, CTCF alone, or both were incubated for 3 min on a dried $NiCl_2$ treated mica-surface before washing off the unbound sample and performing measurements in AC mode. With isolated DNA, nicely separated strands spread on the mica surface were observed **(Figure 4A left)**. Cross-section analysis **(Figure 4A right)** revealed a height of around 2.5 nm, a typical value for DNA in solution (599).

With CTCF, roundish particles of different sizes were observed **(Figure 4B)**. The measured height for the small particles was slightly above 3 nm, which is lower than the assumed diameter of 6.6 nm in the case of CTCF forming a perfect sphere (600) and smaller than the height observed for an 11 ZF construct measured in AFM (222). A possible explanation could be the cantilever tip deforming the protein. Additionally, it is likely that CTCF is not in a globular shape when bound to a surface, as the unstructured termini (29) possibly extend along the surface, which would result in a smaller particle height. However, we were confident to assign the small particles to monomeric CTCF. Cross-sections of the bigger particles revealed a height of more than 10 nm suggesting that these were CTCF oligomers. In agreement with the mass photometry results, this shows that CTCF can form higher oligomeric structures even in the absence of nucleic acids.

Next, CTCF was incubated with DNA and imaged using AFM. Smaller and bigger particles were again observed, which were assigned to monomeric and oligomeric CTCF species **(Figure 4C, left)**. Oligomeric CTCF had DNA incorporated, forming bigger clusters, and it seemed that these clusters contained multiple DNA strands. Cross-sections (red and pink curves) showed variable heights of CTCF-DNA clusters. Note that these clusters differ from CTCF oligomers without DNA as here the centres are flanked by multiple smaller peaks of slightly above 2 nm in case of incorporated DNA and slightly above 3 nm, indicating that monomeric CTCF bound to the outer parts of the CTCF-DNA cluster **(Figure 4C, right)**. CTCF is, therefore, able to form large clusters together with nucleic acids both in cell extracts and by purified CTCF *in vitro*.

### *5.4 Discussion*

CTCF oligomerisation has been observed both *in vitro* (221, 222, 225) and *in vivo* (221, 223, 353). It has thereby been linked to RNA binding (226, 253), DNA network formation and compaction (221, 222) and is known to influence transcription (223, 353) as well as genome architecture (223, 226). Additionally, most conserved topologically associating domain (TAD) boundaries consist of multiple CBSs (601, 602), suggesting a mechanism in which CTCF clustering and crosslinking of DNA could strengthen TAD formation. Here, multiple bulk and single-molecule methods, including mass photometry, bright-field microscopy and AFM, are applied to show that CTCF is able to form oligomers even in the absence of nucleic acids. These oligomers form at physiological salt concentrations and below physiological protein concentrations (597). Using different truncation mutants, it was shown that CTCF's ZFs, but not its unstructured termini, are required for oligomerisation **(Figure 1)**, which is in line with previous results (221, 222).

Furthermore, it was shown that oligomerisation enables CTCF to form mobile droplets at physiological protein and salt concentrations in the presence of a molecular crowder. These droplets were resistant to high-salt and 1,6-hexanediol treatment, which usually resolves droplets formed by LLPS (603). CTCF clusters formed *in vivo* have also been shown to be only modestly

(353) or partly (224) affected by 1,6-hexanediol treatment. As droplets also did not fuse, I assume that they are also not formed by other phase-separation processes like LGPS or PPPS (318). Since CTCF droplets, once formed seem to be highly stable, they might undergo a liquid-to-solid state phase transition, as it has been observed for other proteins (26, 604). However, since CTCF's interaction with chromatin is highly dynamic (605), I believe this to be rather unlikely.

It is, therefore, more likely that the observed droplet formation is not a phase separation process, but light-scattering of large CTCF oligomers. However, this oligomerisation plays an important role as it influences the DNA and RNA interaction of CTCF. AFM experiments **(Figure 4)** revealed that CTCF oligomers and DNA form clusters, similar to a previous AFM study on CTCF's ZFs showing the formation of circular complexes and networks (222). *In vivo,* CTCF and cohesin form colocalised clusters, with CTCF clusters mostly unaffected by cohesin degradation (223), suggesting that CTCF oligomerisation and cluster formation with DNA might enable CTCF to upkeep genomic contacts initially formed by its interaction with cohesin. CTCF's RNA binding, which I showed to depend on CTCF oligomerisation **(Chapter II 4., Figure 5)**, strengthens insulation at TAD boundaries (168, 226, 253) and is involved in SA-dependent cohesin recruitment (261). CTCF's ability to oligomerise might therefore be required for CTCF acting as an anchor point for loop-extruding cohesin complexes and for forming large complexes with DNA even in the absence of cohesin.

While our results suggest that CTCF cannot form condensates on its own *in vitro*, *in vivo* condensate formation might depend on additional cofactors as CTCF was shown to form condensates by colocalisation with the transcriptional coregulator RYBP, which are required for long-range chromatin interactions leading to gene activation (224, 606).

However, it is also likely that CTCF does not form condensates itself but regulates transcription by forming oligomers required for the condensate formation of other proteins. A study has shown that CTCF forms clusters *in vivo* and that loss of these clusters upon CTCF degradation led to an almost complete disappearance of MED1, BRD4 and Pol II condensates (353). These CTCF clusters are hexanediol-stable (353) agreeing with our result that CTCF does not form condensates on its own **(Figure 2)**. Instead, the fact that CTCF can form oligomers on its own and subsequently capture RNA **(Chapter II 4., Figure 5)** might enable interaction with other RBPs including for example BRD4, which is recruited to genomic sites by RNA (345). It could then regulate the positioning of condensates formed by transcription-associated proteins by binding with a high lifetime to its CBS **(Chapter II 4., Figure 1)** and by bringing multiple DNA segments in close contact by the formation of large complexes with DNA **(Figure 4)** thereby facilitating further regulatory interactions of promoter or enhancer bound proteins. A study showed that deletion of a nearby CTCF site led to a loss of super-enhancer-mediated gene expression, which depended on short spatial distances between condensate and gene (349), further suggesting that CTCF oligomerisation might regulate condensate positioning.

CTCF's oligomerisation and the formation of larger complexes with DNA might therefore allow it to regulate the genomic positioning of transcriptional condensates (353) or form condensates on its own in the presence of other cofactors (224). This could provide a model for CTCF's well-established roles in both enabling promoter-enhancer contacts (162, 164–168) within such a condensate or insulating promoter-enhancer contacts (3, 171–173, 175, 176, 179, 181, 203, 607) if either promoter or enhancer is positioned outside of the condensate.

# III Discussion

In this thesis, I have characterised proteins involved in different layers of genome organisation including nucleosome formation, gene regulation and gene expression. This discussion will focus on how the studied proteins influence nuclear processes. The impact of general properties like sequence-specific DNA binding, binding stability or interactions with secondary binding partners will be outlined. In addition, cellular consequences of roadblocking and displacement events between mobile and stationary proteins will be discussed. Furthermore, a model displaying the mutual influence of genome architecture and transcription, depending on oligomerisation and RNA-binding properties of studied protein constructs, is developed.

## III 1. Impact of specific sequence motifs and DNA flexibility on DNA binding

By studying the DNA interactions of different protein constructs presented in this thesis, two different binding modes could be resolved. Either binding to specific target sites, which was found for nucleosomes and CTCF, or sequence-specific movement along the DNA, which was found for cohesin complexes as well as for cohesin's SA subunit, causing an accumulation on AT-rich regions.

These observations represent two general interaction modes of DNA-binding proteins (76): hydrogen bonding with specific base pairs (bps) via the DNA major groove leading to targeting of specific sequence motifs and apart from that interactions of positively charged residues with the DNA minor groove depending on the groove width (76). Preferential binding to narrower minor grooves is caused by stronger electrostatic interactions between positive amino acid residues, most often arginine, and the negatively charged DNA (76). Arginine is a stronger base than lysine with a higher isoelectric point and is therefore positively charged at higher pHs (608) and the guanidinium group of arginine is larger than the amino group of lysine, which means that less energy is required to disrupt the interactions with the surrounding solvent when the amino acid is positioned inside the minor groove (76, 609, 610). In addition, arginine can also form more hydrogen bonds than lysine's amino group in three different directions allowing more conformational freedom and stronger interactions inside the DNA minor groove (76, 611). Groove width influences these interactions and depends on AT content (76) since AT content influences the mechanical properties of DNA.

GC-rich sequences show increased short-range bending, while AT-rich sequences are more prone to melt and to form curved, looped and circular structures on a larger length scale (curvature) (546, 547). The program BEND predicts bending and curvature using common DNA geometry parameters like twist, slide, roll and tilt (612) **(Figure 9A)** based on existing models for gel migration behaviours and structural data of different DNA sequences (546). A model based on nucleosome positioning data thereby predicts a roll that narrows the major groove for GC-rich sequences (positive roll) and a roll that narrows the minor groove for AT-rich sequences (negative roll) (60, 546). Similarly, gel mobility assays display that GC-rich DNA bending towards the major groove and AT-rich DNA bending towards the minor groove will cancel each other out leading to non-curved DNA if separated by roughly one helical turn (613).

Crystal structures also display rolling into the major groove for GC segments (614, 615) and lower minor groove width but no local bending for AT segments (616, 617), suggesting that larger scale curvature stems from intervening G or C nucleotides in AT-rich DNA (618). Interestingly rolling into

the major groove has also been observed for a TA step, which might therefore widen the narrow minor grooves in AT-rich segments (618).

This ability of DNA bps to be changed in their positions or angles with respect to straight B-DNA has been quantified as the TRX score, which evaluates different phosphate backbone angle conformations BI ($\varepsilon$-$\zeta < 0°$) and BII ($\varepsilon$-$\zeta > 0°$) based on NMR and crystallographic structures (544, 545). The TRX score is defined as the average BII population of the two opposite phosphates of each dinucleotide segment (544, 545). AT-rich segments contain overall more BI conformations, in which bps are constrained in their positioning with regard to the helical axis (low shift and slide), display almost no bending (close to zero roll), and no large deviations from B-DNA twist angles, leading to rigid DNA structures with narrow minor grooves (544, 545). In contrast, GC-rich segments often contain alternating BI and BII conformations with BII conformations allowing bps to be positioned towards or bend into the major groove (544, 545) **(Figure 9B, top),** leading to shallower major grooves and wider minor grooves (544) **(Figure 9B, bottom)** and explaining higher flexibility of GC-rich regions (544, 547). These sequence properties regarding groove width and flexibility (76, 544, 545) could therefore help to understand some of the here studied protein-DNA interactions.

A typical DNA-binding motif that targets AT-rich regions is the AT-hook, which contains an Arg-Gly-Arg core forming electrostatic interactions with the minor groove (619–622), which might be preferential for narrower minor grooves (76). The AT-hook motif enables SA1 to target telomeric sequences but is absent in SA2 (623). However, our study (229) revealed AT-rich binding for both SA1 and SA2.

Both consist mainly of HEAT-repeat domains, required for the interaction with Rad21, WAPL, CTCF and DNA therefore influencing both loop-anchoring and DNA-unbinding of cohesin (18, 183, 200, 212, 214, 624, 625). The cohesin loader subunit Scc2/NIPBL also consists mainly of HEAT-repeats which form interactions with the cohesin ring and DNA (99, 626). HEAT-repeats are composed of parallel stacks of two antiparallel amphiphilic α-helices linked by a short loop each, that form flexible, curved, superhelical or C/U-shaped structures enclosing DNA by forming electrostatic interactions with the DNA backbone (183, 627–630). SA1 binds to DNA through multiple lysines and arginines in an extended loop and an α-helix (183). The *yeast* ortholog Scc3 forms a positive surface with the kleisin Scc1 targeting the phosphate backbone via multiple lysines (625), which in both cases might favour narrow minor DNA grooves (76). SA2 contains multiple positively charged surfaces that preferentially target DNA secondary structures (214, 435), and most DNA-binding interfaces of NIPBL are unspecific interactions with the phosphate backbone via positively charged lysine and arginines (183), which could explain the absence of specific sequence motifs for all HEAT-repeat cohesin cofactors and instead favour AT-rich binding (183, 214, 435, 625). Further examples can be found in other SMC complexes like condensin. The condensin HEAT-repeat subunit YCG1 contains a DNA-binding cleft (631), much like SA1, suggesting a paradigmatic role of HEAT-repeat-mediated DNA-binding for SMC complexes (625, 632). HEAT-repeat proteins like the cohesin loader or Scc3 (SA in *humans*) binding preferentially to AT-rich regions could therefore cause AT-rich binding of the cohesin complex *in vivo*, observed for example in *Saccharomyces cerevisiae* (369, 633). Additionally, cohesin preferentially binds to AT-rich promoters before it is translocated by transcribing polymerases (634) and the loading of cohesin onto centromeric sequences (188, 189, 635), could be caused by AT-rich centromeric DNA (636). The single-molecule experiments performed in **Chapter II 3.** and **Chapter 4.** show that cohesin and SA target AT-rich regions *in vitro*, in line with a previous study (197), which showed AT-rich binding for *S. pombe* cohesin and the cohesin loader, suggesting that the two HEAT-repeat cohesin cofactors SA and the cohesin loader cause AT-rich binding of cohesin.

Another group of minor-groove binders are high mobility group (HMG) box proteins (637–639). Electrostatic and hydrophobic interactions with the minor groove are mediated by an L-shaped

structure formed by three α-helices, which upon binding cause a widening of the minor groove resulting in DNA bending of up to 110° (638, 640, 641). HMG box proteins display a low lifetime on cellular DNA (642), which by cycles of unbinding and rebinding could allow them to quickly scan the DNA for secondary DNA structures appearing during DNA replication and DNA repair (637–640, 643). In our study (229), SA also displayed a low lifetime at physiological salt concentrations and quantitative photobleaching analysis revealed that short-lived DNA binding can be found among many DNA-binding proteins in the nucleus (642). Additionally, I showed (229) that SA enriches on AT-rich regions by faster diffusion on GC-rich regions. AT-rich sequence targeting by diffusion has also been shown for cohesin (197) and this diffusive behaviour might enable cohesin to accumulate on DNA-repair sites (241, 644), by moving along the DNA until reaching sites of DNA damage recognised by SA2 (435).

Further examples of proteins scanning the DNA are sliding repair glycosylases and hopping/sliding Endonuclease V, which share a similar mechanism of being able to push a wedge motif between DNA strands at DNA lesions during diffusion along the DNA allowing them to slow down at these DNA damage sites (645–649). Another example is the DNA polymerase processivity factor UL42, which binds the DNA backbone by electrostatic interactions and moves along the DNA by hopping while being able to wait for polymerase arrival by an increased lifetime on DNA ends (650, 651).

Taken together weaker electrostatic interactions with wider minor grooves in GC-rich regions seem to lead to fast diffusion and shorter binding to DNA. This might enable proteins to quickly scan chromosomal DNA for other protein binding partners or sites of DNA damage.

Besides the recognition of AT-rich DNA regions, proteins can also recognise specific DNA sequences. An important sequence-specific DNA-binding motif are zinc fingers (ZFs), which recognise three-bp sequences and are found in many DNA-binding proteins involved in transcriptional regulation and DNA repair (207, 652, 653).

ZFs can be divided into many different folding groups (207). Interestingly different DNA interactions can be observed for the same class of ZFs, like for the Zn2/Cys6-like fingers (207, 654). While most family members like Gal4 and PPR1 bind spaced GC-rich triplets by hydrogen bonds with the major groove, PUT3 additionally widens the minor groove by inserting a β-strand resulting in DNA bending of a sequence between two bound GC-rich triplets (654). Similarly, while most treble clef fingers interact with the major groove, the endonuclease I-TevI interacts with the minor groove and phosphate backbone, displaying less sequence-specific binding (207, 655, 656) again showing different interaction modes for the same class of ZFs.

The C2H2 ZF motif is a common feature of many TFs and DNA-binding proteins, which target specific sequence motifs (207). A computational tool has been developed to predict position weight matrices (PWMs) of different C2H2 ZFs, including a PWM for ZFs 2-11 of CTCF (657). PWMs display the DNA positions of a protein target site on the x-axis and show the probability of each of the four DNA bases occurring at each given position on the y-axis. Interestingly, the simulation data PWM displays a good agreement with *in vivo* data for the core motif but a less good prediction for the CTCF upstream motif (204, 657). Strikingly, in (229) I show that a mutant containing only the upstream motif-binding ZFs 9-11 is enriched on AT-rich sequences on DNA. CTCF's outer ZFs are presumed to bind to a sequence-specific upstream motif (204), by forming hydrogen bonds between arginine and glutamine residues with conserved bps of this motif (206). However, these three ZFs contain a total of 40 arginine and lysine residues and might show a preference for narrow DNA minor grooves by additional electrostatic interactions with the minor groove (76). Our study (229) shows that these electrostatic interactions outweigh sequence-specific hydrogen-bond formation since CTCF's outer ZFs did not recognise the upstream motif in our λ-DNA construct in the absence of core motif-binding ZFs (229). The existence of specific upstream motifs *in vivo* (204) can have different causes than ZF sequence specificity, like interaction with other binding partners

or DNA occupancy by other proteins. In contrast, full-length CTCF displayed sequence-specific binding in our study (229). It was enriched on CTCF-binding sites (CBSs) by an increased lifetime and salt stability on these CBSs compared to unspecific DNA (229). This reveals different binding modes, sequence-motif-specific major groove binding as well as binding to AT-rich sequences, existing for the same class of ZFs and even within the same molecule.

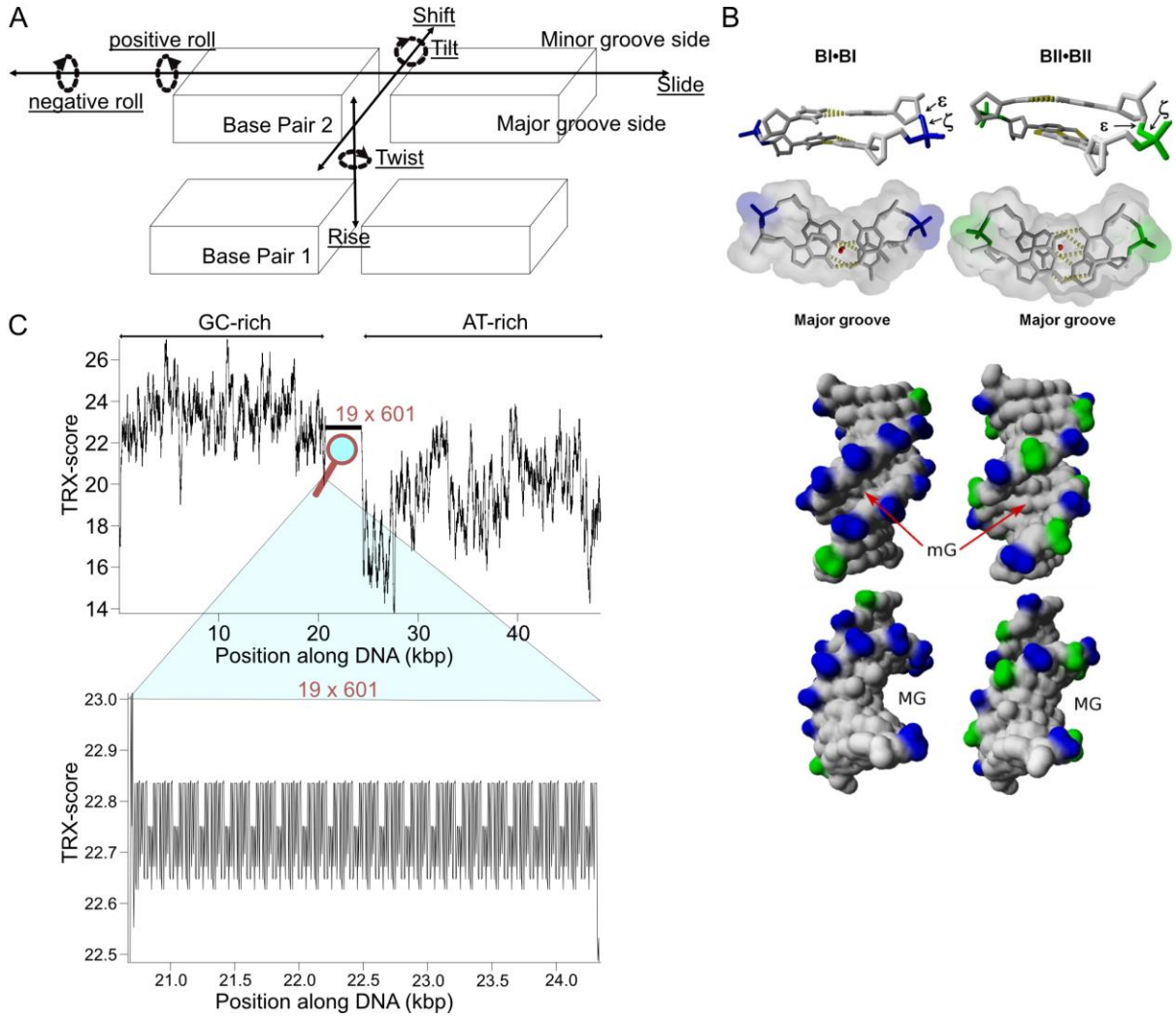



*Figure 9: AT-ratio of λ-DNA influences its flexibility, which can be quantified by the TRX score. A) DNA geometry is based on 6 parameters. In B-DNA the twist angle (rotation around the helical axis) is around 34.3° (10.5 bps for a 360° turn) and the step size between bps is around 0.34 nm (rise). Bps can rotate around their bond (roll) or around an axis perpendicular to their bond and the helical axis (tilt). Positive rolls narrow the major and negative rolls the minor groove. Tilting causes the rise between the two consecutive bps to differ for the two DNA strands. Additionally, the bases can be displaced along the direction of their bond (slide) or perpendicular to it (shift). Figure based on (658). B) Top: BI steps (in blue) differ in phosphate backbone torsion angles (ε-ζ < 0°) from BII steps (ε-ζ > 0°) (in green). For BI segments bps are positioned close to the centre of the helix (red) leading to a deeper major groove compared to BII segments, for which bps are positioned further towards (shift) and bend into the major groove (roll) (544, 545). Bottom: BI-rich DNA (left) contains a narrower minor groove (mG), leading to stronger electrostatic interactions with amino acid side chains, and a deeper major groove (MG) compared to the BII-rich DNA (right), with BI segments coloured in blue and BII segments coloured in green (76, 544). X-Ray structures: PDB: 1EHV, left and PDB: 3GGI, right (544). Figure from (544). C) Top: The λ-DNA construct contains a GC-rich region (high TRX scores) and an AT-rich region (low TRX scores). The 19 x 601 sites in between*

*display a periodicity between high and low TRX scores. (545) Bottom: Zoomed curve of the 19 x 601 sites.*

In our study, CTCF enriched on CBSs displayed a higher lifetime (229) than measured previously (12, 201) and remained stably bound to its CBS at even above physiological salt concentrations. This high stability of CTCF at topologically associating domain (TAD) boundaries enables CTCF to form long-lived DNA-loops together with cohesin (215). However static binding with a high lifetime might impede target-site search. This is known as the speed-stability paradox (659, 660). A model dependent on two different binding modes (specific vs. unspecific with different dissociation rates) shows that proteins can overcome the speed-stability problem if unbinding, 3D-diffusion and rebinding are accompanied by movement along the DNA (1D-diffusion) (659). TFs display different mechanisms to find their target site efficiently (661–663). This does include lower lifetimes on unspecific sequences allowing unbinding and rebinding until a specific sequence is found (3D-search) (664–668), which is however slow and can on its own not explain the fast target search of many TFs (662, 669). It can also include 1D diffusion along unspecific DNA, which can be either sliding along the DNA or moving in small jumps of dissociation and reassociation to the DNA (hopping) (664, 665, 670–673). Furthermore, transient bridging of two DNA segments followed by rebinding to the second segment (intersegment transfer) enables fast scanning of different DNA segments as well as overcoming roadblocks (674–679). Finally, the formation of protein/nucleic acid clusters leads to faster DNA association by high local concentrations of the TF, enables intersegment transfer via protein oligomerisation and increases the chance of rebinding a target site within a cluster after dissociation (662, 666, 677, 680, 681).

In (229) we show that CTCF molecules can perform target-site search by significantly faster 1D-diffusion on non-CBSs, compared to its relatively static binding to CBSs. This can be explained by two different binding modes with a different number of formed hydrogen bonds, since hydrogen bond formation depends on the interaction with specific bases along the CBS inside the DNA major groove (30). Alternatively, transient unbinding of single or multiple ZFs from unspecific sequences would reduce the binding stability resulting in a significantly lowered lifetime and faster diffusion on non-CBSs. The existence of CTCF motifs without the upstream motif (204) implies that unbinding of outer ZFs could be a frequent phenomenon *in vivo*. This could also enable CTCF intersegment transfer, which has been observed for another C2H2 ZF Egr-1 (660, 674). Egr-1 can change into a search mode, in which the least stably bound of three ZF unbinds from the first DNA segment and then rebinds to transiently bridge two DNA molecules before Egr-1 continues its search on the second DNA segment (674). Additionally, I showed that CTCF forms oligomeric structures that form clusters with RNA **(Chapter II 4.)** and DNA **(Chapter II 5.)**. This oligomerisation can enhance target-site search, since single particle tracking combined with simulations has shown that CTCF diffusion *in vivo* is restricted to certain regions, possibly caused by RNA-mediated clustering, increasing the CBS association rate, while preventing localisation to non-CBS regions (680). Therefore the formation of multiple hydrogen bonds with the DNA major groove allows for a long lifetime on specific target sites, but additional DNA-binding properties, like clustering, intersegment transfer by the existence of multiple binding sites or lower lifetime and faster 1D diffusion on unspecific sites are required for fast target-site search.

Similar to CTCF, histone octamers recognise specific DNA sequences. Nucleosomes showed a clear enrichment on the 601 sites **(Chapter II 2., Figure 3B,C)**, which contain a 10 bp periodicity between high and low TRX scores **(Figure 9C)** (544, 545). Regions of high TRX scores at positions where the major groove faces towards the octamer allow for stronger bending of DNA and base pair shifting towards the major groove, therefore reducing the energetic costs of nucleosome wrapping (544, 545). AT-dinucleotides causing lower TRX-scores in between increase the energetic gain from electrostatic interactions at positions of inwards pointing minor grooves with histone octamer arginine side chains (10, 50, 76). Reduction of bending energy by GC-rich

sequences is the more crucial factor as **Chapter II 2** shows that nucleosomes not only target a specific sequence motif but additionally favour GC-rich over AT-rich sequences, in contrast to other proteins like SA or cohesin, which are shown in this thesis to favour AT-rich sequences. A study (58) displayed that nucleosome occupancy *in vivo* increases until a certain GC content from where it starts to decline again, which can be explained by the large loss of entropy when the DNA is tightly wrapped around the histone octamer for sequences with very high GC content (545). For example at the +1 nucleosome in *mice*, the maximum for nucleosome occupancy was found to be slightly below 60 % GC content (58) a value close to many GC-rich regions and 601 sites in my single-molecule measurements **(Chapter II 2., Figure 3B),** which displayed higher nucleosome occupancy than AT-rich regions.

The binding of other DNA-bending proteins has also been shown to depend on GC content. For example, the GC-rich target sequence of the TF PU.1 adopts a bend conformation that allows for stronger electrostatic interactions, with TF binding leading to an even stronger DNA bending (682). The transcriptional regulator bovine papillomavirus E2 protein binds two major grooves of a GC-rich recognition site by two α-helices with additional interactions being formed with the phosphate backbone by bending the DNA in between around a β-barrel (683). A strikingly different example of DNA bending is the TATA-box-binding protein TBP which binds the AT-rich TATA-box and causes DNA melting by inducing an 80°C bend (684–687). Thus, DNA bending might be either facilitated by the higher bendability of GC-rich regions (544–547) or by a reduced energetic cost of DNA melting for AT-rich sequences (547, 688, 689).

Low nucleosome occupancy levels can be found in AT-rich regions in *yeast* around TSSs (54, 56) but also on CpG islands (51, 67). CpG islands are GC-rich sequences often found near mammalian promoters and enriched in TF-binding sites (68) and in CBSs (73). These CpG islands form multiple long-range contacts with each other, suggesting that CTCF binding to CpG islands causes DNA cluster formation involved in CTCF's role in genome organisation (73). On the one hand, nucleosome instability on CpG islands and increased DNA accessibility (51, 67) might, therefore, allow increased TF binding (68), since TFs display faster association rates and higher lifetimes when DNA is not covered by nucleosomes (690). On the other hand, TFs binding to GC-rich cis-regulatory elements (652) and CpG islands (68–72) could also prevent nucleosome formation (68). This could be caused by a specific class of TFs called pioneer transcription factors, which can enhance DNA accessibility by interacting with histones or nucleosome-bound DNA weakening histone-DNA interactions or contacts between nucleosomes (149, 569, 570, 578). TFs preferentially binding to GC-rich regions and CpG islands might be caused by the strong interactions between arginines and guanines (691, 692) as observed for CTCF forming multiple hydrogen bonds with arginines of ZFs 3-7 and the guanines of the core motif (30, 206).

In summary, there are many structural and functional differences for proteins targeting specific DNA sequence motifs. Sequence-specific hydrogen bonds with the DNA major groove like for CTCF's ZFs (30) lead to static and highly stable DNA binding. This allows the formation of long-lived complexes but might also impede target-site search due to slower sampling or diffusion depending on the stability of the protein on unspecific sites (659). In contrast, AT-rich binders, like cohesin's SA subunit, display a more diffusive and less stable DNA-binding behaviour, which results in faster scanning for protein binding partners, AT-rich sequences or secondary DNA/RNA structures but prevents targeting of specific sequence motifs and formation of long-lived entities on DNA. This fast scanning can enable SA and/or cohesin to interact with different structures along the DNA, which could include R-loops, CTCF, telomers, or DNA repair structures (216, 228, 435, 623). Finally, nucleosome assembly on GC-rich regions or regions with alternating GC content depends on both the strength of electrostatic interactions and DNA bendability (10, 63, 76, 545). As discussed for ZFs, these different binding modes do not seem to be restricted to certain protein families (207, 654).

## III 2. Revealing interactions of colliding molecules on DNA curtains

In this chapter, I take a closer look at specific protein-protein interactions happening on the DNA track: Collision of diffusing proteins or molecular motors with statically DNA-bound proteins like CTCF or paused polymerases, are frequent but not well-understood scenarios, which can influence nuclear processes like transcription, replication and TAD formation (247, 365, 404, 693, 694).

As there are 55,000 to 65,000 CBSs in mammalian genomes (695), collisions of translocating or diffusing proteins with static CBS-bound CTCFs are frequent nuclear events. Our study (229) displayed that diffusing CTCF molecules were mostly blocked and continued their diffusive behaviour on the same site of a CBS-bound CTCF. Rarely CTCF was able to pass the roadblock but was unable to push CTCF off its CBS or stay bound to the CBS-bound CTCF. This is contradicting to CTCF oligomerisation in solution **(Chapter II 5.)**. Since oligomerisation was dependent on CTCF's ZFs, which are accessible and not saturated in solution, DNA-bound CTCF might be inhibited in oligomerisation due to the occupied ZFs.

For SA most diffusing molecules interacted with and stayed bound to the CTCF roadblock (229). This might be a mechanism of how SA can interact with CTCF *in vivo*, independently of cohesin (216) and lead to direct SA-mediated loading of cohesin at CBSs. Furthermore, diffusing SAs were mostly blocked by CTCF and bypass happened only rarely. This demonstrates that roadblocks can indeed constrict diffusion and lead to SA and possibly cohesin being confined in their diffusion to CBS-defined regions. However, the observed bypassing events indicate that DNA-bound proteins can be imperfect barriers and that additional binding partners can be required for a complete blockage at TAD boundaries. Recently, it was reported that CTCF acts as an imperfect barrier and that blockage was direction- and tension-dependent (12). This raises questions about the actual mechanism behind TAD insulation (18, 19, 40) and indicates that boundary-associated RNAs and transcription can impact TAD insulation (166, 168, 226, 253, 261) by, for example, increasing roadblocking efficiency through the formation of larger complexes and steric blocking.

However, not all DNA-bound molecules can be confined by roadblocks, which would drastically impede most nuclear processes. Possible mechanisms are hopping over protein obstacles, like DNA repair glycosylases (696), or intersegment transfer observed, for example, for argonaute proteins (697). Argonaute proteins perform 1D diffusion on DNA for target-site search but can also jump to DNA segments close in 3D space and thereby overcome obstacles (697), displaying a different intersegment transfer mechanism than explained above for Erg1, which uses multiple binding sites (674). The nucleotide excision repair protein Rad23B displayed salt-dependent 1D diffusion on DNA by hopping short distances and was able to pass over small obstacles by jumping over them (481). Similar to SA proteins, XPC-Rad23B enriched on AT-rich regions by slowed diffusion at these regions (481). Slowed diffusion at partially melted AT-rich regions (698–700) together with the ability for obstacles passage can allow the fast search for DNA-damage sites (481) and could be a common mechanism for XPC-Rad23B, SA and other proteins involved in DNA repair.

The ring-shaped mismatch repair protein Mlh1–Pms1 complexes can move past each other and past obstacles larger than their ring size, which requires switching between a closed and an open ring configuration (479, 701). Cohesin complexes also form ring-like structures, which can bind DNA in its lumen (191). This topologically bound cohesin can overcome small roadblocks fitting through its large ring size, like nucleosomes or restriction enzymes, by diffusion (197). This suggests sliding over obstacles as a bypass mechanism (197) instead of ring-opening. In addition, cohesin was found to bypass also very large obstacle during loop extrusion (198), which suggests a different, non-topological binding mode for cohesin and a controlled DNA handover of different binding sites to achieve processive movement (702). This implies that loop extrusion is another

possibility to overcome obstacles, which is a processive movement in 3D space along the DNA with large step sizes (12). Structural features like forming large or transiently opened rings around DNA, the presence of multiple DNA-binding domains allowing intersegment transfer, or jumping over obstacles by dissociation and subsequent rebinding can therefore enable diffusing molecules to overcome roadblocks.

Besides diffusing molecules, the encounter of roadblocks is a major challenge for molecular motors moving directionally on the DNA track, where obstacle bypass is not possible. These scenarios happen frequently in transcription elongation, DNA replication or DNA repair (693). The bacterial helicase and nuclease RecBCD can efficiently clear the DNA track by pushing molecules like lac repressor, stalled RNA polymerases, nucleosomes, or EcoRI$^{E111Q}$ over long distances along the DNA before causing them to unbind (693, 703), which could allow it to quickly translocate to and then start unwinding DNA at both potentially lethal double-strand breaks (DSBs) (704, 705). In contrast, the bacterial translocase FtsK displayed different collision behaviours, including pausing and direction reversal, but also passing, presumably by ring-opening, and roadblock removal stimulated by FtsK cooperativity (487). During bacterial replication, the helicase Rep undergoes cycles of unbinding and rebinding from the replisome and removes roadblocks in an ATPase-dependent manner including R-loops formed by inactivated Cas9, which in the absence of Rep inhibit DNA replication (694). Similarly in eukaryotic replication, the accessory helicase Pif1 was shown to be required for replication to proceed past inactivated Cas9 roadblocks, by removing both RNA and Cas9 (706). As R-loops are common byproducts of transcription these accessory helicases therefore play important roles in resolving transcription-replication conflicts in multiple organisms. These examples show that clearing the DNA of roadblocks seems to be a common mechanism for motor proteins (693), which increases DNA accessibility for DNA downstream processes in replication and repair.

Apart from replication and DNA-repair roadblocks also impair transcription processes. DNA-bound roadblocks often cause early termination, which can have important regulatory functions like preventing read-through into a second gene or protecting weakly expressed genomic regions (406, 417). RNA polymerases can also be forced to wait for roadblock dissociation or in the presence of GreA undergo cycles of backtracking, GreA-stimulated transcript cleavage and restart until they are able to cause roadblock unbinding (405). Additionally to the presence of cofactors, cooperativity is also often required for bacterial RNA polymerases. A second trailing polymerase can cause the restart of a paused or backtracked leading polymerase and help it overcome the roadblock (402, 409, 514).

Eukaryotic RNA polymerases frequently need to bypass nucleosomes on their DNA track and a mechanism has been proposed that includes transient pausing while weakening histone dimer-DNA interactions before DNA-loop formation leads to the new assembly of the nucleosome behind the polymerase (374, 392). The passage might include H2A/H2B dimer unbinding (393, 398) and require the activity of CRCs weakening histone dimer-DNA interactions (400) as *in vitro* polymerases display pausing when encountering nucleosomes depending on the stability of the nucleosome on the bound DNA (395). This suggests that polymerases might wait for partial unbinding of the DNA from the H2A/H2B dimer (395). RNA polymerases therefore often pause, stop and backtrack when encountering roadblocks and might require help from cofactors or additional polymerases.

In **Chapter II 4.** I show that both CTCF alone and CTCF-SA complexes are pushed off by transcribing T7 polymerases, although both are very stably bound to the DNA. The speed of the polymerases is not or only marginally changed during encounters. This shows that transcribing T7 polymerases are efficient at removing protein roadblocks, similar to the above-mentioned examples in replication and repair. Previously, T7 bulk experiments suggested that during polymerase collisions in most cases, a trailing polymerase can push off a leading polymerase by “collision-

induced bumping", but only after the leading polymerase has cleared the promoter (535). In contrast, I showed that in most cases, the trailing polymerase would pause after a collision. However, the leading polymerase did not always remain at the pause site but was in some cases, also removed from the DNA. In fewer cases, the stalled leading polymerase was pushed along the DNA by the trailing polymerase, with both polymerases continuing to move along the DNA. This displays a mechanism of how multiple polymerases starting from the same promoter can clear genes of stalled complexes. This is important for efficient transcription during the T7 infection cycle, since after T7 polymerase and T7 lysozyme expression other genes required for virus assembly are subsequently transcribed by the T7 polymerase (491–494) and T7 polymerase is required for primer formation for viral genome replication (513, 707).

In summary, roadblocks impede transcription by causing stalling, backtracking or termination of transcribing polymerases (404), which can be overcome by the cooperativity of polymerases or by the presence of additional cofactors (402, 403, 405, 409).

In conclusion, DNA-binding proteins, as described for CTCF, act as roadblocks for diffusing proteins, confining their 1D target-site search to a constricted region. However mechanisms to overcome roadblocks have been shown including transient ring-opening (479, 487, 701), hopping (696) or intersegment transfer (697). In addition, motor proteins can efficiently remove obstacles from DNA, enabling DNA repair or replication (693). For transcribing bacterial polymerases, synergistic action is required to overcome roadblocks (402, 409) and a second T7 polymerase can push paused complexes off or along the DNA (this thesis), enabling efficient transcription.

## III 3. Transcriptional regulation is shaped by protein oligomerisation

After describing DNA interactions on the level of single molecules, this chapter will focus on how higher-order complex formation impacts DNA-binding behaviour. Protein homo-oligomerisation plays an important role in many cellular processes (708, 709). For example, it is critical for the function of membrane-pore-forming proteins (710–712). This often relies on hydrophobic interactions (713) or on interactions between unstructured and elongated protein segments called intrinsically disordered regions (IDRs) (714). In the nucleus, homo-oligomerisation is often found among DNA binders and is required for targeting specific sequence motifs (715) or forming protein-DNA networks (716).

Consequently, homo-oligomerisation contributes to different transcription regulation processes. For example, the herpesvirus protein LANA oligomerises on viral terminal repeats forming a higher-order DNA complex with many binding sites bridged into one large network, leading to transcriptional repression and therefore autoregulation of LANA (717). The transcription repressor Proline-Rich Homeodomain protein (PRH/Hex) is able to form oligomeric complexes in solution, which are then able to form compacted DNA fibres on DNA segments containing multiple PRH-binding sites (718). Both cases show that homo-oligomerisation enables DNA-binding proteins to drive distant DNA sites in close proximity. This establishes homo-oligomerisation as an alternative or at least an additional mechanism of DNA loop formation for the establishment of TADs. For CTCF, homo-oligomerisation led to the formation of large complexes with DNA **(Chapter II 5)**. This would analogously bring distance genomic sites into close contact, independently of cohesin, which can be confirmed by *in vivo* observations of CTCF clusters remaining stable after Smc3 degradation (223). The formation of higher order-complexes by CTCF may enable CTCF to perform its role as a transcriptional insulator by spatially separating certain genomic regions from each other (3, 171–173, 175, 176, 179, 181, 203, 607) or enable the formation of promoter-enhancer contacts within these large complexes (162, 164–169).

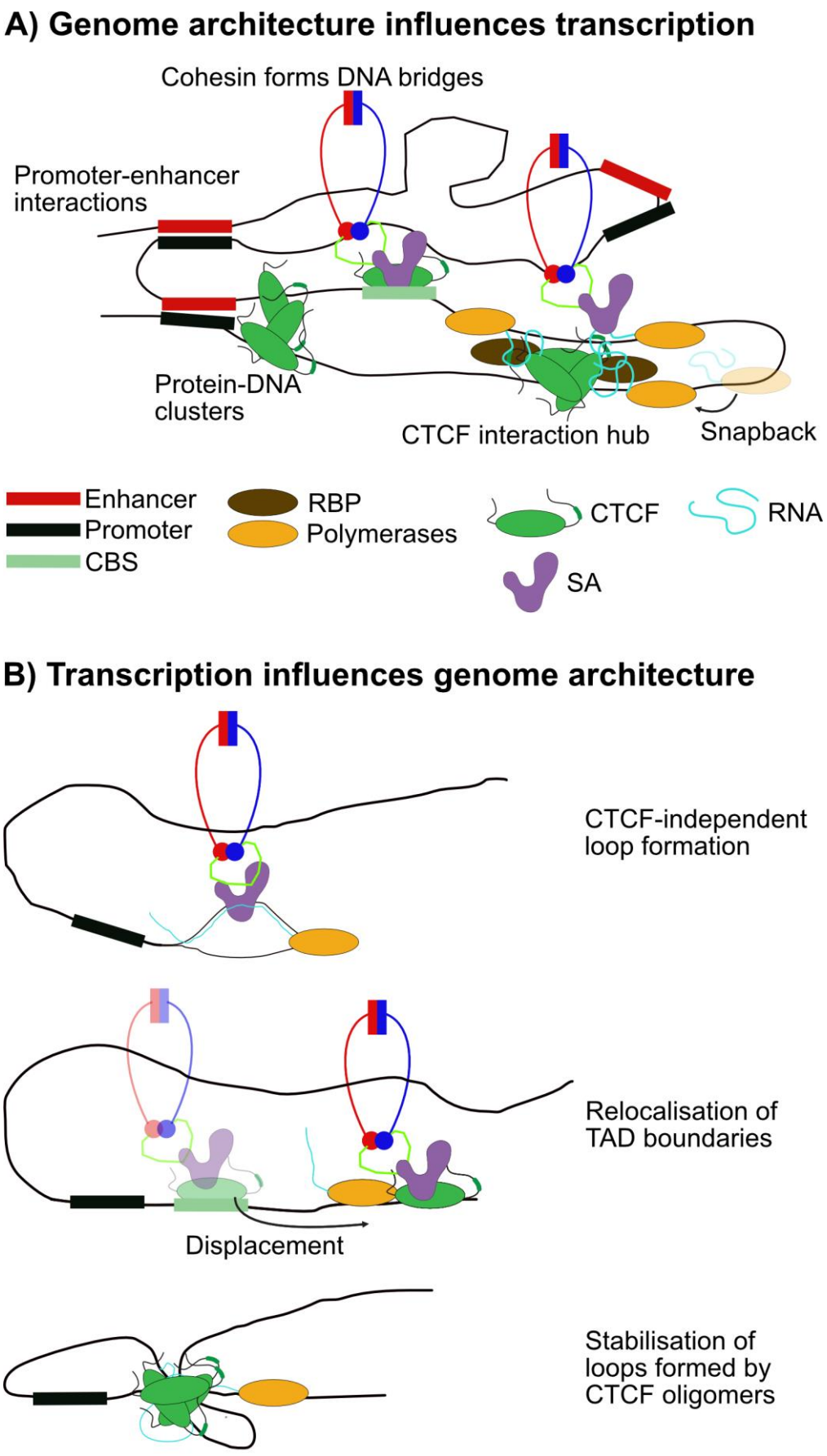



*Figure 10: Model of the interplay between genome architecture and transcription. A) Transcription regulation mediated by cohesin and CTCF. Cohesin gets recruited to CTCF-bound sites via a direct interaction between its SA subunit and CTCF's ZFs or via SA binding to CTCF-bound RNA.*

*Promoter-enhancer interaction frequency could be increased by cohesin-mediated DNA bridges* (230) *and by DNA/RNA-protein interaction hubs formed by CTCF oligomerisation. CTCF oligomers can then interact with RNAs and recruit other RBPs. Additionally, CTCF can locally increase polymerase concentration by causing an increased number of polymerase pauses and snapbacks contributing to the formation of clusters of enhanced transcription. B) Independently of CTCF, SA binding to RNA structures can enable transcription-regulated TAD formation. Transcribing polymerases can displace SA and CTCF, which would affect the positions of TAD boundaries. CTCF oligomers were stabilised on unspecific positions by RNA, which could lead to the formation of DNA/RNA-protein clusters at sites of active transcription.*

Additionally, I showed (229) that homo-oligomerisation allows CTCF to perform secondary RNA capture while already being bound to DNA. CTCF clustering and interaction with RNA can enhance CTCF target-site search (680) by restricting diffusion to confined spaces and help CTCF overcome the above-mentioned speed-stability paradox (659) and oligomerisation by itself can facilitate target-site search by enabling intersegment transfer (677). Additionally, by enabling RNA binding, oligomerisation allows CTCF to interact with different nuclear RNAs including ncRNAs, which is required for forming strong TAD boundaries (168, 226, 253) and can facilitate cohesin recruitment by CTCF (261). RNA-capture via CTCF oligomerisation is therefore critical for CTCF finding its CBS and for forming strong anchor points for TAD formation.

Besides cluster formation, phase separation is another phenomenon leading to the assembly of large structures tightly associated with nuclear regulatory processes (27). The *human* single-stranded DNA binding protein (hSSB), required for multiple DNA repair pathways and genome stability, forms condensates via an IDR and a nucleic acid-binding domain preferentially with single-stranded DNA (ssDNA) (719, 720). Besides DNA repair, condensate formation is associated with transcriptional regulation. Many transcriptional regulators contain IDRs and a model has been proposed in which super-enhancers activate genes by the formation of large condensates enriched in components of the transcription machinery (25, 313, 319). They are characterized by the regulation of cell type-specific genes and burst expression patterns since the positioning of a gene within or outside of a condensate leads to quick gene activation or inactivation respectively (313, 347, 349, 350, 721).

The phase-separation abilities of these nuclear proteins depend on IDRs, nucleic acid-binding domains or both. As CTCF contains nucleic acid-binding domains (30) as well as IDRs (29), it might be involved in nuclear phase separation and condensation processes. However, the droplet experiments **(Chapter II 5.)** suggest that CTCF is not able to perform phase separation on its own. A recent study showed that CTCF clusters are required for BRD4, MED1 and RNA Pol II transcriptional condensate formation *in vivo* (353), which are all three involved in the formation of super-enhancer condensates (25, 313, 319). The here detected high stability on CBSs **(Chapter II 4.)** and its ability to form oligomers on its own **(Chapter II 5.),** which could then form clusters with DNA and RNA, could explain how CTCF can act as this observed start point for the formation of transcriptional condensates (353). The positioning of CBSs can then determine the localisation of CTCF clusters and possibly of super-enhancer formation.

I therefore propose that CTCF oligomerisation plays an important role in transcription regulation **(Figure 10A)**. Oligomerisation, RNA capture **(Chapter II 4.)**, and formation of larger complexes with DNA **(Chapter II 5.)** by CTCF could provide a platform for the recruitment of other RNA-binding proteins involved in the regulation of gene expression like the above-mentioned BRD-4, which is recruited by enhancer-RNAs (345). I observed (229) that oligomerisation and subsequent RNA capture stabilise CTCF even on unspecific binding positions. Transcription-produced RNA might therefore allow CTCF to stably bind to low-affinity CBSs *in vivo*, which seem to function mostly in transcriptional regulation (33), in contrast to high-affinity CBSs, which are involved in genome architecture by regulating heterochromatin spreading and being more often colocalised with

cohesin (33, 722). Our transcription experiments revealed that CTCF increases the frequency of polymerase pausing and snapbacks (229) which could lead to a local increase in polymerase concentrations within these interaction hubs. CTCF can recruit SA to these interaction hubs (216, 229) and SA-mediated cohesin loading can then lead to the formation of DNA loops by bridging DNA in cis as displayed in **Chapter II 3.** (230), enabling further promoter-enhancer interactions.

## III 4. Transcription influences genome architecture

After outlining how architectural proteins can regulate transcription, the influence of transcribing polymerases on genome architecture will be described for different organisms in this chapter (163). In *Saccharomyces cerevisiae,* where CTCF is not present, TAD boundaries are associated with enhanced transcription (723) and polymerases can push cohesin along the DNA (367, 634), suggesting that transcription can replace the architectural role of CTCF in *yeast*.

In *human* and *mouse* genomes, a subset of TAD boundaries overlap with factors associated with active promoters instead of CBSs (16) and transcribing polymerases were shown to be able to take over the role of CTCF in *mice* acting as moving TAD boundaries (366), showing that transcription is sufficient for blocking cohesin loop extrusion even in organisms expressing CTCF.

In *Drosophila*, inter-TAD regions are commonly found at positions of accessible and strongly transcribed chromatin and TAD boundaries at positions where architectural protein binding sites colocalise with promoters of active genes (44, 363, 364). Therefore TAD boundaries in *Drosophila* are mainly regulated by transcription (44, 363), while CTCF boundary function is only required for a small subset of *Drosophila* TADs (362, 364).

CTCF-independent TAD formation could be explained by preferential loading of cohesin to sites of active transcription by the presence of R-loops (216). Our DNA curtain experiments (229) reveal that cohesin's SA subunit has a higher lifetime on secondary RNA structures than on DNA. In line, SA preferentially binds to secondary RNA structures in AFM measurements (228) and colocalises with R-loops *in vivo* (216), which could lead to cohesin complex loading at sites of active transcription **(Figure 10B).**

Additionally to TAD formation, transcription also influences TAD positioning. Transcription elongation by RNA polymerase II can push cohesin *in vivo* (366), reposition it to active genes (368) and remove cohesin and in some cases also CTCF from CBSs (370). This is in line with our DNA curtains experiments in which CTCF, as well as CTCF-SA complexes, were pushed off CBSs by transcribing T7 polymerases (229) **(Figure 10B)**. This raises the question of how CTCF can shape genome architecture after being displaced by transcription. *In vivo*, CTCF's association with chromatin is highly dynamic (605), and CTCF's high nuclear concentration (597) might lead to fast reoccupation of CBSs after CTCF displacement. This might be required at some lower affinity CBSs, which function in regulating gene expression in actively transcribed regions, and display a low CTCF occupancy after CTCF depletion (33) suggesting that CTCF might be pushed off these sites by transcribing polymerases.

Contrary results were reported for the influence of transcription inhibition on CTCF chromatin associations. Transcription inhibition can lead to increased CTCF cluster formation, depending on the presence of cohesin (223), again suggesting that CTCF and cohesin loop anchors could be disturbed by transcription. In contrast, transcription inhibition was shown to disrupt CTCF-chromatin interactions, most significantly at promoters and TSSs, suggesting that CTCF is stabilised by transcription-produced RNAs (253). In line, I observed that CTCF oligomerisation and subsequent RNA capture stabilise CTCF even on unspecific DNA-binding positions (229). Therefore, there

could be different kinds of CTCF-dependent TAD formation in the nucleus. The first is formed by CTCF acting as a boundary for cohesin loop extrusion and is disturbed by transcription (366, 368, 370). The second type is formed by CTCF oligomerisation and stabilised by transcription-produced RNAs (168, 253) **(Figure 10B)**. In line two different kinds of chromatin loops exist *in vivo* with only one requiring CTCF's RNA binding function (226).

In summary, I propose that transcription influences genome architecture in multiple ways **(Figure 10B)**. Transcription could lead to the recruitment of cohesin to RNA-secondary structures via its SA subunit and to CTCF-independent loop formation. Additionally, transcription could displace CTCF or CTCF interacting with SA or the whole cohesin complex from CBSs leading to changes in TAD structure. Finally, transcription could stabilise DNA loops formed by CTCF oligomers independently of cohesin by enabling CTCF to capture transcription-produced RNA.

## III 5. Conclusion

The creation of nucleosome assemblies, single-molecule transcription assays and site-specific roadblock formation on DNA curtains allowed me to study genome architecture proteins and their interactions with polymerases. The results reveal that there is a mutual interplay between transcriptional regulation and genome architecture and that CTCF's and cohesin's roles in these processes are more complex than CTCF merely acting as an anchor point for loop extruding cohesin complexes.

CTCF bound its CBS statically and with a high lifetime (229) enabling it to form stable anchor points at TAD boundaries for DNA loops formed together with cohesin (18, 19), where it may be able to recruit cohesin via our observed direct interaction with cohesin's SA subunit (229). Transcribing RNA polymerases were able to push CTCF and CTCF-SA off CBSs (229), revealing a mechanism of how TAD boundaries and cohesin-dependent CTCF clusters can be disturbed by transcription (223, 370). CTCF's ability to oligomerise and form large clusters with DNA suggests that genomic contacts can form within DNA-CTCF clusters independently of cohesin. Oligomerised CTCF was able to capture RNA (229), which led to a stabilisation of CTCF even on non-CBSs and could create interaction hubs for additional proteins. These interaction hubs could consequently be stabilised by the RNA from close-by transcription (33, 168, 253) and allow CTCF to perform its known functions of regulating transcription by enabling regulatory contacts within these interaction hubs (162, 164–168) while preventing interactions with promoters or enhancers outside of them (3, 170–173, 175, 176, 181, 203). Furthermore, this could allow CTCF to influence transcription by recruiting additional RNAs and proteins involved in transcriptional regulation (166, 216, 224, 260, 261, 724–728) and by causing the formation of transcriptional condensates (353). Our DNA curtains assay with CBS-enriched CTCF and CTCF-RNA clusters can be applied to study if CTCF can recruit transcriptional regulators and super-enhancer-forming proteins to these sites.

In contrast to the stably bound CTCF, cohesin's SA subunit preferentially targeted AT-rich regions in a diffusive transient manner. Unlike CTCF, SA might not be able to create an interaction hub for other proteins. Instead, this shows that SA can quickly scan the DNA and could then be stabilised at TAD boundaries by CTCF (216, 229), at DNA repair intermediates (435), or at RNA structures (216, 228, 229). I propose that SA then loads cohesin to these sites, enabling cohesin's roles in DNA repair (232, 240, 241) and directing TAD boundaries to both sites of active genes and CBSs (16). This suggests that SA is the actual cohesin loader in higher eukaryotes, which could also help explain why SA1-cohesin and SA2-cohesin display distinct genomic binding sites, recruiting cohesin mostly to CBSs or sites of active transcription respectively (186, 187). Our study showed (230) that the *yeast* cohesin loader Scc2/4 is not required for topological DNA entrapment. Scc2/4

stimulated cohesin's ATPase activity (230) and was required for stable entrapment of two DNA segments, presumably in two different cohesin ring compartments (230). This suggests that Scc2/4 main roles are in facilitating cohesin's roles in generating DNA loops for TAD formation and entrapment of two DNAs for sister chromatid cohesion and possibly as already shown for the *human* loader (14) in enabling cohesin loop extrusion. DNA curtain measurements performed for SA could be repeated for the *yeast* cohesin cofactors Scc2/4 and Scc3 to study their enrichment and possible cohesin loading function on for example AT-rich sequences, RNA-transcripts and DNA repair intermediates.

The studied T7 RNA polymerase was found to push other polymerases as well as genome architecture proteins along the DNA track. This leads to more efficient transcription and also shows that transcription is not only influenced by but also shapes 3D genome organisation.

# IV References


1. Piovesan,A., Pelleri,M.C., Antonaros,F., Strippoli,P., Caracausi,M. and Vitale,L. (2019) On the length, weight and GC content of the human genome. *BMC Res. Notes*, **12**, 106.

2. Pal,K., Forcato,M. and Ferrari,F. (2019) Hi-C analysis: from data generation to integration. *Biophys. Rev.*, **11**, 67–78.

3. 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.

4. Eeftens,J. and Dekker,C. (2017) Catching DNA with hoops—biophysical approaches to clarify the mechanism of SMC proteins. *Nat. Struct. Mol. Biol.*, **24**, 1012–1020.

5. Lai,W.K.M. and Pugh,B.F. (2017) Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.*, **18**, 548–562.

6. Liu,R., Wu,J., Guo,H., Yao,W., Li,S., Lu,Y., Jia,Y., Liang,X., Tang,J. and Zhang,H. (2023) Post□ translational modifications of histones: Mechanisms, biological functions, and therapeutic targets. *MedComm*, **4**, e292.

7. Monteagudo-Sánchez,A., Richard Albert,J., Scarpa,M., Noordermeer,D. and Greenberg,M.V.C. (2024) The impact of the embryonic DNA methylation program on CTCF-mediated genome regulation. *Nucleic Acids Res.*, **52**, 10934–10950.

8. 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.*Vol. 472, pp. 293–315.

9. Silverman,B. and Mirsky,A.E. (1973) Accessibility of DNA in Chromatin to DNA Polymerase and RNA Polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 1326–1330.

10. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.

11. Prieto,E.I. and Maeshima,K. (2019) Dynamic chromatin organization in the cell. *Essays Biochem.*, **63**, 133–145.

12. 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.

13. Fudenberg,G., Imakaev,M., Lu,C., Goloborodko,A., Abdennur,N. and Mirny,L.A. (2016) Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.*, **15**, 2038–2049.

14. Davidson,I.F., Bauer,B., Goetz,D., Tang,W., Wutz,G. and Peters,J.-M. (2019) DNA loop extrusion by human cohesin. *Science.*, **366**, 1338–1345.

15. Kim,Y., Shi,Z., Zhang,H., Finkelstein,I.J. and Yu,H. (2019) Human cohesin compacts DNA by loop extrusion. *Science.*, **366**, 1345–1349.

16. 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.

17. Sexton,T., Yaffe,E., Kenigsberg,E., Bantignies,F., Leblanc,B., Hoichman,M., Parrinello,H., Tanay,A. and Cavalli,G. (2012) Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*, **148**, 458–472.

18. 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.

19. 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.e22.

20. Michaelis,C., Ciosk,R. and Nasmyth,K. (1997) Cohesins: Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids. *Cell*, **91**, 35–45.

21. Lieberman-Aiden,E., van Berkum,N.L., 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**, 289–293.

22. Larson,A.G., Elnatan,D., Keenen,M.M., Trnka,M.J., Johnston,J.B., Burlingame,A.L., Agard,D.A., Redding,S. and Narlikar,G.J. (2017) Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, **547**, 236–240.

23. Strom,A.R., Emelyanov,A. V., Mir,M., Fyodorov,D. V., Darzacq,X. and Karpen,G.H. (2017) Phase separation drives heterochromatin domain formation. *Nature*, **547**, 241–245.

24. Misteli,T. (2020) The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*, **183**, 28–45.

25. Cho,W.-K., Spille,J.-H., Hecht,M., Lee,C., Li,C., Grube,V. and Cisse,I.I. (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science.*, **361**, 412–415.

26. Boeynaems,S., Alberti,S., Fawzi,N.L., Mittag,T., Polymenidou,M., Rousseau,F., Schymkowitz,J., Shorter,J., Wolozin,B., Van Den Bosch,L., *et al.* (2018) Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.*, **28**, 420–435.

27. Feric,M. and Misteli,T. (2021) Phase separation in genome organization across evolution. *Trends Cell Biol.*, **31**, 671–685.

28. Wiedner,H.J. and Giudice,J. (2021) It's not just a phase: function and characteristics of RNA-binding proteins in phase separation. *Nat. Struct. Mol. Biol.*, **28**, 465–473.

29. Martinez,S.R. and Miranda,J.L. (2010) CTCF terminal segments are unstructured. *Protein Sci.*, **19**, 1110–1116.

30. Yang,J., Horton,J.R., Liu,B., Corces,V.G., Blumenthal,R.M., Zhang,X. and Cheng,X. (2023) Structures of CTCF–DNA complexes including all 11 zinc fingers. *Nucleic Acids Res.*, **51**, 8447–8462.

31. Bar-Ziv,R., Voichek,Y. and Barkai,N. (2016) Chromatin dynamics during DNA replication. *Genome Res.*, **26**, 1245–1256.

32. Hilbert,L., Sato,Y., Kuznetsova,K., Bianucci,T., Kimura,H., Jülicher,F., Honigmann,A., Zaburdaev,V. and Vastenhouw,N.L. (2021) Transcription organizes euchromatin via microphase separation. *Nat. Commun.*, **12**, 1360.

33. 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.

34. van Attikum,H., Fritsch,O., Hohn,B. and Gasser,S.M. (2004) Recruitment of the INO80 Complex by H2A Phosphorylation Links ATP-Dependent Chromatin Remodeling with DNA Double-Strand Break Repair. *Cell*, **119**, 777–788.

35. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing Chromosome Conformation. *Science.*, **295**, 1306–1311.

36. Belton,J.-M., McCord,R.P., Gibcus,J.H., Naumova,N., Zhan,Y. and Dekker,J. (2012) Hi–C: A comprehensive technique to capture the conformation of genomes. *Methods*, **58**, 268–276.

37. Magaña-Acosta,M. and Valadez-Graham,V. (2020) Chromatin Remodelers in the 3D Nuclear Compartment. *Front. Genet.*, **11**, 600615.

38. Cremer,T. and Cremer,M. (2010) Chromosome Territories. *Cold Spring Harb. Perspect. Biol.*, **2**, a003889.

39. Sikorska,N. and Sexton,T. (2020) Defining Functionally Relevant Spatial Chromatin Domains: It is a TAD Complicated. *J. Mol. Biol.*, **432**, 653–664.

40. 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.

41. Guelen,L., Pagie,L., Brasset,E., Meuleman,W., Faza,M.B., Talhout,W., Eussen,B.H., de Klein,A., Wessels,L., de Laat,W., *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948–951.

42. van Steensel,B. and Belmont,A.S. (2017) Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*, **169**, 780–791.

43. Kind,J., Pagie,L., de Vries,S.S., Nahidiazar,L., Dey,S.S., Bienko,M., Zhan,Y., Lajoie,B., de Graaf,C.A., Amendola,M., *et al.* (2015) Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. *Cell*, **163**, 134–147.

44. Rowley,M.J., Nichols,M.H., Lyu,X., Ando-Kuri,M., Rivera,I.S.M., Hermetz,K., Wang,P., Ruan,Y. and Corces,V.G. (2017) Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell*, **67**, 837-852.e7.

45. Olins,A.L. and Olins,D.E. (1974) Spheroid Chromatin Units (v Bodies). *Science.*, **183**, 330–332.

46. Kornberg,R. (1974) Chromatin Structure : A Repeating Unit of Histones and DNA. *Science.*, **184**, 868–871.

47. Luger,K. and Richmond,T.J. (1998) The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.*, **8**, 140–146.

48. Li,W., Dou,S.-X. and Wang,P.-Y. (2005) The histone octamer influences the wrapping direction of DNA on it: Brownian dynamics simulation of the nucleosome chirality. *J. Theor. Biol.*, **235**, 365–372.

49. Bruno,M., Flaus,A., Stockdale,C., Rencurel,C., Ferreira,H. and Owen-Hughes,T. (2003) Histone H2A/H2B Dimer Exchange by ATP-Dependent Chromatin Remodeling Activities. *Mol. Cell*, **12**, 1599–1606.

50. Drew,H.R. and Travers,A.A. (1985) DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.*, **186**, 773–790.

51. Valouev,A., Johnson,S.M., Boyd,S.D., Smith,C.L., Fire,A.Z. and Sidow,A. (2011) Determinants of nucleosome organization in primary human cells. *Nature*, **474**, 516–520.

52. Schones,D.E., Cui,K., Cuddapah,S., Roh,T.-Y., Barski,A., Wang,Z., Wei,G. and Zhao,K. (2008) Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, **132**, 887–898.

53. Valouev,A., Ichikawa,J., Tonthat,T., Stuart,J., Ranade,S., Peckham,H., Zeng,K., Malek,J.A., Costa,G., McKernan,K., *et al.* (2008) A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Res.*, **18**, 1051–1063.

54. Yuan,G.-C., Liu,Y.-J., Dion,M.F., Slack,M.D., Wu,L.F., Altschuler,S.J. and Rando,O.J. (2005) Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. *Science.*, **309**, 626–630.

55. Szerlong,H.J. and Hansen,J.C. (2011) Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem. Cell Biol.*, **89**, 24–34.

56. Kaplan,N., Moore,I.K., Fondufe-Mittendorf,Y., Gossett,A.J., Tillo,D., Field,Y., LeProust,E.M., Hughes,T.R., Lieb,J.D., Widom,J., *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, **458**, 362–366.

57. Tillo,D., Kaplan,N., Moore,I.K., Fondufe-Mittendorf,Y., Gossett,A.J., Field,Y., Lieb,J.D., Widom,J., Segal,E. and Hughes,T.R. (2010) High Nucleosome Occupancy Is Encoded at Human Regulatory Sequences. *PLoS One*, **5**, e9129.

58. Fenouil,R., Cauchy,P., Koch,F., Descostes,N., Cabeza,J.Z., Innocenti,C., Ferrier,P., Spicuglia,S., Gut,M., Gut,I., *et al.* (2012) CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.*, **22**, 2399–2408.

59. Krietenstein,N., Wal,M., Watanabe,S., Park,B., Peterson,C.L., Pugh,B.F. and Korber,P. (2016)

Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell*, **167**, 709-721.e12.

60. Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.*, **191**, 659–675.

61. Segal,E., Fondufe-Mittendorf,Y., Chen,L., Thåström,A., Field,Y., Moore,I.K., Wang,J.-P.Z. and Widom,J. (2006) A genomic code for nucleosome positioning. *Nature*, **442**, 772–778.

62. Lowman,H. and Bina,M. (1990) Correlation between dinucleotide periodicities and nucleosome positioning on mouse satellite DNA. *Biopolymers*, **30**, 861–876.

63. Lowary,P.T. and Widom,J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, **276**, 19–42.

64. Anderson,J.D. and Widom,J. (2001) Poly(dA-dT) Promoter Elements Increase the Equilibrium Accessibility of Nucleosomal DNA Target Sites. *Mol. Cell. Biol.*, **21**, 3830–3839.

65. Kaplan,N., Moore,I., Fondufe-Mittendorf,Y., Gossett,A.J., Tillo,D., Field,Y., Hughes,T.R., Lieb,J.D., Widom,J. and Segal,E. (2010) Nucleosome sequence preferences influence in vivo nucleosome organization. *Nat. Struct. Mol. Biol.*, **17**, 918–920.

66. Sekinger,E.A., Moqtaderi,Z. and Struhl,K. (2005) Intrinsic Histone-DNA Interactions and Low Nucleosome Density Are Important for Preferential Accessibility of Promoter Regions in Yeast. *Mol. Cell*, **18**, 735–748.

67. Ramirez-Carrozzi,V.R., Braas,D., Bhatt,D.M., Cheng,C.S., Hong,C., Doty,K.R., Black,J.C., Hoffmann,A., Carey,M. and Smale,S.T. (2009) A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. *Cell*, **138**, 114–128.

68. Deaton,A.M. and Bird,A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.

69. Rozenberg,J.M., Shlyakhtenko,A., Glass,K., Rishi,V., Myakishev,M. V, FitzGerald,P.C. and Vinson,C. (2008) All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. *BMC Genomics*, **9**, 67.

70. Jaeger,S.A., Chan,E.T., Berger,M.F., Stottmann,R., Hughes,T.R. and Bulyk,M.L. (2010) Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. *Genomics*, **95**, 185–195.

71. Landolin,J.M., Johnson,D.S., Trinklein,N.D., Aldred,S.F., Medina,C., Shulha,H., Weng,Z. and Myers,R.M. (2010) Sequence features that drive human promoter function and tissue specificity. *Genome Res.*, **20**, 890–898.

72. Gebhard,C., Benner,C., Ehrich,M., Schwarzfischer,L., Schilling,E., Klug,M., Dietmaier,W., Thiede,C., Holler,E., Andreesen,R., *et al.* (2010) General Transcription Factor Binding at CpG Islands in Normal Cells Correlates with Resistance to De novo DNA Methylation in Cancer Cells. *Cancer Res.*, **70**, 1398–1407.

73. Gushchanskaya,E.S., Artemov,A. V, Ulyanov,S. V, Logacheva,M.D., Penin,A.A., Kotova,E.S., Akopov,S.B., Nikolaev,L.G., Iarovaia,O. V, Sverdlov,E.D., *et al.* (2014) The clustering of CpG islands may constitute an important determinant of the 3D organization of interphase chromosomes. *Epigenetics*, **9**, 951–963.

74. Ichikawa,Y., Morohashi,N., Tomita,N., Mitchell,A.P., Kurumizaka,H. and Shimizu,M. (2016) Sequence-directed nucleosome-depletion is sufficient to activate transcription from a yeast core promoter in vivo. *Biochem. Biophys. Res. Commun.*, **476**, 57–62.

75. Handy,D.E., Castro,R. and Loscalzo,J. (2011) Epigenetic Modifications: Basic Mechanisms and Role in Cardiovascular Disease. *Circulation*, **123**, 2145–2156.

76. 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.

77. Li,S., Peng,Y., Landsman,D. and Panchenko,A.R. (2022) DNA methylation cues in nucleosome geometry, stability and unwrapping. *Nucleic Acids Res.*, **50**, 1864–1874.

78. Lin,J.C., Jeong,S., Liang,G., Takai,D., Fatemi,M., Tsai,Y.C., Egger,G., Gal-Yam,E.N. and Jones,P.A. (2007) Role of Nucleosomal Occupancy in the Epigenetic Silencing of the MLH1 CpG Island. *Cancer Cell*, **12**, 432–444.

79. Portela,A., Liz,J., Nogales,V., Setién,F., Villanueva,A. and Esteller,M. (2013) DNA methylation determines nucleosome occupancy in the 5′-CpG islands of tumor suppressor genes. *Oncogene*, **32**, 5421–5428.

80. Lee,J.Y. and Lee,T.-H. (2012) Effects of DNA Methylation on the Structure of Nucleosomes. *J. Am. Chem. Soc.*, **134**, 173–175.

81. Cutter,A.R. and Hayes,J.J. (2015) A brief review of nucleosome structure. *FEBS Lett.*, **589**, 2914–2922.

82. Clark,M.A., Douglas,M. and Choi,J. (2018) 16.3 Eukaryotic epigenetic gene regulation. In *Biology 2e*. OpenStax, Houston, Texas.

83. Allahverdi,A., Yang,R., Korolev,N., Fan,Y., Davey,C.A., Liu,C.-F. and Nordenskiöld,L. (2011) The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Res.*, **39**, 1680–1691.

84. Ausio,J., Dong,F. and van Holde,K.E. (1989) Use of selectively trypsinized nucleosome core particles to analyze the role of the histone “tails” in the stabilization of the nucleosome. *J. Mol. Biol.*, **206**, 451–463.

85. Sokolova,V., Miratsky,J., Svetlov,V., Brenowitz,M., Vant,J., Lewis,T.S., Dryden,K., Lee,G., Sarkar,S., Nudler,E., *et al.* (2024) Structural mechanism of HP1α-dependent transcriptional repression and chromatin compaction. *Structure.*, **32**, 2094-2106.e6.

86. Hong,L., Schroth,G.P., Matthews,H.R., Yau,P. and Bradbury,E.M. (1993) Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 ‘tail’ to DNA. *J. Biol. Chem.*, **268**, 305–14.

87. Shogren-Knaak,M., Ishii,H., Sun,J.-M., Pazin,M.J., Davie,J.R. and Peterson,C.L. (2006) Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science.*, **311**, 844–847.

88. Narita,T., Higashijima,Y., Kilic,S., Liebner,T., Walter,J. and Choudhary,C. (2023) Acetylation of histone H2B marks active enhancers and predicts CBP/p300 target genes. *Nat. Genet.*, **55**, 679–692.

89. Esmaeili,M., Blythe,S.A., Tobias,J.W., Zhang,K., Yang,J. and Klein,P.S. (2020) Chromatin accessibility and histone acetylation in the regulation of competence in early development. *Dev. Biol.*, **462**, 20–35.

90. Hebbes,T.R., Thorne,A.W. and Crane-Robinson,C. (1988) A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.*, **7**, 1395–1402.

91. Brownell,J.E., Zhou,J., Ranalli,T., Kobayashi,R., Edmondson,D.G., Roth,S.Y. and Allis,C.D. (1996) Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation. *Cell*, **84**, 843–851.

92. Lee,D.Y., Hayes,J.J., Pruss,D. and Wolffe,A.P. (1993) A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*, **72**, 73–84.

93. Biernat,E., Kinney,J., Dunlap,K., Rizza,C. and Govind,C.K. (2021) The RSC complex remodels nucleosomes in transcribed coding sequences and promotes transcription in Saccharomyces cerevisiae. *Genetics*, **217**, iyab021.

94. Agalioti,T., Chen,G. and Thanos,D. (2002) Deciphering the Transcriptional Histone Acetylation Code for a Human Gene. *Cell*, **111**, 381–392.

95. Parnell,T.J., Huff,J.T. and Cairns,B.R. (2008) RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. *EMBO J.*, **27**, 100–110.

96. Sims,R.J., Chen,C.-F., Santos-Rosa,H., Kouzarides,T., Patel,S.S. and Reinberg,D. (2005) Human but Not Yeast CHD1 Binds Directly and Selectively to Histone H3 Methylated at Lysine 4 via Its Tandem Chromodomains. *J. Biol. Chem.*, **280**, 41789–41792.

97. Lauberth,S.M., Nakayama,T., Wu,X., Ferris,A.L., Tang,Z., Hughes,S.H. and Roeder,R.G. (2013) H3K4me3 Interactions with TAF3 Regulate Preinitiation Complex Assembly and Selective Gene Activation. *Cell*, **152**, 1021–1036.

98. Flanagan,J.F., Mi,L.-Z., Chruszcz,M., Cymborowski,M., Clines,K.L., Kim,Y., Minor,W., Rastinejad,F. and Khorasanizadeh,S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, **438**, 1181–1185.

99. Park,S., Kim,G.W., Kwon,S.H. and Lee,J. (2020) Broad domains of histone H3 lysine 4 trimethylation in transcriptional regulation and disease. *FEBS J.*, **287**, 2891–2902.

100. Wang,T., Xu,C., Liu,Y., Fan,K., Li,Z., Sun,X., Ouyang,H., Zhang,X., Zhang,J., Li,Y., *et al.* (2012) Crystal Structure of the Human SUV39H1 Chromodomain and Its Recognition of Histone H3K9me2/3. *PLoS One*, **7**, e52977.

101. Peters,A.H.F.M., O'Carroll,D., Scherthan,H., Mechtler,K., Sauer,S., Schöfer,C., Weipoltshammer,K., Pagani,M., Lachner,M., Kohlmaier,A., *et al.* (2001) Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell*, **107**, 323–337.

102. Cheutin,T., McNairn,A.J., Jenuwein,T., Gilbert,D.M., Singh,P.B. and Misteli,T. (2003) Maintenance of Stable Heterochromatin Domains by Dynamic HP1 Binding. *Science.*, **299**, 721–725.

103. Nakayama,J., Rice,J.C., Strahl,B.D., Allis,C.D. and Grewal,S.I.S. (2001) Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science.*, **292**, 110–113.

104. Lachner,M., O'Carroll,D., Rea,S., Mechtler,K. and Jenuwein,T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, **410**, 116–120.

105. Hiragami-Hamada,K., Soeroes,S., Nikolov,M., Wilkins,B., Kreuz,S., Chen,C., De La Rosa-Velázquez,I.A., Zenn,H.M., Kost,N., Pohl,W., *et al.* (2016) Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. *Nat. Commun.*, **7**, 11310.

106. Bulut-Karslioglu,A., De La Rosa-Velázquez,I.A., Ramirez,F., Barenboim,M., Onishi-Seebacher,M., Arand,J., Galán,C., Winter,G.E., Engist,B., Gerle,B., *et al.* (2014) Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Mol. Cell*, **55**, 277–290.

107. Todd,L.A., Hall,A.C., Pietrobon,V., Chan,J.N.Y., Laflamme,G. and Mekhail,K. (2020) RNA-cDNA hybrids mediate transposition via different mechanisms. *Sci. Rep.*, **10**, 16034.

108. Zeller,P., Padeken,J., van Schendel,R., Kalck,V., Tijsterman,M. and Gasser,S.M. (2016) Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat. Genet.*, **48**, 1385–1395.

109. Allshire,R.C. and Madhani,H.D. (2018) Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.*, **19**, 229–244.

110. Nonaka,N., Kitajima,T., Yokobayashi,S., Xiao,G., Yamamoto,M., Grewal,S.I.S. and Watanabe,Y. (2002) Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nat. Cell Biol.*, **4**, 89–93.

111. Gregan,J., Riedel,C.G., Pidoux,A.L., Katou,Y., Rumpf,C., Schleiffer,A., Kearsey,S.E., Shirahige,K., Allshire,R.C. and Nasmyth,K. (2007) The Kinetochore Proteins Pcs1 and Mde4 and Heterochromatin Are Required to Prevent Merotelic Orientation. *Curr. Biol.*, **17**, 1190–1200.

112. 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.

113. Valgardsdottir,R., Chiodi,I., Giordano,M., Rossi,A., Bazzini,S., Ghigna,C., Riva,S. and Biamonti,G. (2008) Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res.*, **36**, 423–434.

114. Eymery,A., Horard,B., Atifi-Borel,M. El, Fourel,G., Berger,F., Vitte,A.-L., Van den Broeck,A.,

Brambilla,E., Fournier,A., Callanan,M., *et al.* (2009) A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. *Nucleic Acids Res.*, **37**, 6340–6354.

115. Saksouk,N., Simboeck,E. and Déjardin,J. (2015) Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin*, **8**, 3.

116. Malik,H.S. and Henikoff,S. (2003) Phylogenomics of the nucleosome. *Nat. Struct. Mol. Biol.*, **10**, 882–891.

117. West,M.H. and Bonner,W.M. (1980) Histone 2A, a heteromorphous family of eight protein species. *Biochemistry*, **19**, 3238–3245.

118. Bönisch,C. and Hake,S.B. (2012) Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic Acids Res.*, **40**, 10719–10741.

119. Mizuguchi,G., Shen,X., Landry,J., Wu,W.-H., Sen,S. and Wu,C. (2004) ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. *Science.*, **303**, 343–348.

120. Willhoft,O., Ghoneim,M., Lin,C.-L., Chua,E.Y.D., Wilkinson,M., Chaban,Y., Ayala,R., McCormack,E.A., Ocloo,L., Rueda,D.S., *et al.* (2018) Structure and dynamics of the yeast SWR1-nucleosome complex. *Science.*, **362**, eaat7716.

121. Heo,K., Kim,H., Choi,S.H., Choi,J., Kim,K., Gu,J., Lieber,M.R., Yang,A.S. and An,W. (2008) FACT-Mediated Exchange of Histone Variant H2AX Regulated by Phosphorylation of H2AX and ADP-Ribosylation of Spt16. *Mol. Cell*, **30**, 86–97.

122. Rogakou,E.P., Pilch,D.R., Orr,A.H., Ivanova,V.S. and Bonner,W.M. (1998) DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *J. Biol. Chem.*, **273**, 5858–5868.

123. Siino,J.S., Nazarov,I.B., Svetlova,M.P., Solovjeva,L. V, Adamson,R.H., Zalenskaya,I.A., Yau,P.M., Bradbury,E.M. and Tomilin,N. V (2002) Photobleaching of GFP-labeled H2AX in chromatin: H2AX has low diffusional mobility in the nucleus. *Biochem. Biophys. Res. Commun.*, **297**, 1318–1323.

124. Paull,T.T., Rogakou,E.P., Yamazaki,V., Kirchgessner,C.U., Gellert,M. and Bonner,W.M. (2000) A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.*, **10**, 886–895.

125. Fernandez-Capetillo,O., Lee,A., Nussenzweig,M. and Nussenzweig,A. (2004) H2AX: the histone guardian of the genome. *DNA Repair (Amst).*, **3**, 959–967.

126. Xu,Y., Ayrapetov,M.K., Xu,C., Gursoy-Yuzugullu,O., Hu,Y. and Price,B.D. (2012) Histone H2A.Z Controls a Critical Chromatin Remodeling Step Required for DNA Double-Strand Break Repair. *Mol. Cell*, **48**, 723–733.

127. Brunelle,M., Nordell Markovits,A., Rodrigue,S., Lupien,M., Jacques,P.-É. and Gévry,N. (2015) The histone variant H2A.Z is an important regulator of enhancer activity. *Nucleic Acids Res.*, **43**, 9742–9756.

128. Hu,G., Cui,K., Northrup,D., Liu,C., Wang,C., Tang,Q., Ge,K., Levens,D., Crane-Robinson,C. and Zhao,K. (2013) H2A.Z Facilitates Access of Active and Repressive Complexes to Chromatin in Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell*, **12**, 180–192.

129. Bernstein,B.E., Mikkelsen,T.S., Xie,X., Kamal,M., Huebert,D.J., Cuff,J., Fry,B., Meissner,A., Wernig,M., Plath,K., *et al.* (2006) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, **125**, 315–326.

130. Allfrey,V.G., Faulkner,R. and Mirsky,A.E. (1964) ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc. Natl. Acad. Sci. U.S.A.*, **51**, 786–794.

131. Li,F., Martienssen,R. and Cande,W.Z. (2011) Coordination of DNA replication and histone modification by the Rik1–Dos2 complex. *Nature*, **475**, 244–248.

132. Furuta,T., Takemura,H., Liao,Z.-Y., Aune,G.J., Redon,C., Sedelnikova,O.A., Pilch,D.R., Rogakou,E.P., Celeste,A., Chen,H.T., *et al.* (2003) Phosphorylation of Histone H2AX and

Activation of Mre11, Rad50, and Nbs1 in Response to Replication-dependent DNA Double-strand Breaks Induced by Mammalian DNA Topoisomerase I Cleavage Complexes. *J. Biol. Chem.*, **278**, 20303–20312.

133. Buontempo,S., Laise,P., Hughes,J.M., Trattaro,S., Das,V., Rencurel,C. and Testa,G. (2022) EZH2-Mediated H3K27me3 Targets Transcriptional Circuits of Neuronal Differentiation. *Front. Neurosci.*, **16**, 814144.

134. Weber,C.M., Ramachandran,S. and Henikoff,S. (2014) Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase. *Mol. Cell*, **53**, 819–830.

135. Azmi,I.F., Watanabe,S., Maloney,M.F., Kang,S., Belsky,J.A., MacAlpine,D.M., Peterson,C.L. and Bell,S.P. (2017) Nucleosomes influence multiple steps during replication initiation. *Elife*, **6**, e22512.

136. Clapier,C.R. and Cairns,B.R. (2009) The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem.*, **78**, 273–304.

137. Eustermann,S., Patel,A.B., Hopfner,K.-P., He,Y. and Korber,P. (2024) Energy-driven genome regulation by ATP-dependent chromatin remodellers. *Nat. Rev. Mol. Cell Biol.*, **25**, 309–332.

138. Hargreaves,D.C. and Crabtree,G.R. (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.*, **21**, 396–420.

139. Eustermann,S., Schall,K., Kostrewa,D., Lakomek,K., Strauss,M., Moldt,M. and Hopfner,K. (2018) Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature*, **556**, 386–390.

140. Zhou,C.Y., Johnson,S.L., Lee,L.J., Longhurst,A.D., Beckwith,S.L., Johnson,M.J., Morrison,A.J. and Narlikar,G.J. (2018) The Yeast INO80 Complex Operates as a Tunable DNA Length-Sensitive Switch to Regulate Nucleosome Sliding. *Mol. Cell*, **69**, 677-688.e9.

141. Rawal,Y., Chereji,R. V, Qiu,H., Ananthakrishnan,S., Govind,C.K., Clark,D.J. and Hinnebusch,A.G. (2018) SWI/SNF and RSC cooperate to reposition and evict promoter nucleosomes at highly expressed genes in yeast. *Genes Dev.*, **32**, 695–710.

142. Alatwi,H.E. and Downs,J.A. (2015) Removal of H2A.Z by INO80 promotes homologous recombination. *EMBO Rep.*, **16**, 986–994.

143. Oberbeckmann,E., Krietenstein,N., Niebauer,V., Wang,Y., Schall,K., Moldt,M., Straub,T., Rohs,R., Hopfner,K.-P., Korber,P., *et al.* (2021) Genome information processing by the INO80 chromatin remodeler positions nucleosomes. *Nat. Commun.*, **12**, 3231.

144. Oberbeckmann,E., Niebauer,V., Watanabe,S., Farnung,L., Moldt,M., Schmid,A., Cramer,P., Peterson,C.L., Eustermann,S., Hopfner,K.-P., *et al.* (2021) Ruler elements in chromatin remodelers set nucleosome array spacing and phasing. *Nat. Commun.*, **12**, 3232.

145. Mao,C., Brown,C.R., Griesenbeck,J. and Boeger,H. (2011) Occlusion of Regulatory Sequences by Promoter Nucleosomes In Vivo. *PLoS One*, **6**, e17521.

146. Ho,L., Miller,E.L., Ronan,J.L., Ho,W.Q., Jothi,R. and Crabtree,G.R. (2011) esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat. Cell Biol.*, **13**, 903–913.

147. Kubik,S., O'Duibhir,E., de Jonge,W.J., Mattarocci,S., Albert,B., Falcone,J.-L., Bruzzone,M.J., Holstege,F.C.P. and Shore,D. (2018) Sequence-Directed Action of RSC Remodeler and General Regulatory Factors Modulates +1 Nucleosome Position to Facilitate Transcription. *Mol. Cell*, **71**, 89-102.e5.

148. Hu,G., Schones,D.E., Cui,K., Ybarra,R., Northrup,D., Tang,Q., Gattinoni,L., Restifo,N.P., Huang,S. and Zhao,K. (2011) Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res.*, **21**, 1650–1658.

149. Mivelaz,M., Cao,A.-M., Kubik,S., Zencir,S., Hovius,R., Boichenko,I., Stachowicz,A.M., Kurat,C.F., Shore,D. and Fierz,B. (2020) Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. *Mol. Cell*, **77**, 488-500.e9.

150. Song,S., Perez,J. V, Svitko,W., Ricketts,M.D., Dean,E., Schultz,D., Marmorstein,R. and

Johnson,F.B. (2020) Rap1-mediated nucleosome displacement can regulate gene expression in senescent cells without impacting the pace of senescence. *Aging Cell*, **19**, e13061.

151. Kubik,S., Bruzzone,M.J., Jacquet,P., Falcone,J.-L., Rougemont,J. and Shore,D. (2015) Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. *Mol. Cell*, **60**, 422–434.

152. Eaton,M.L., Galani,K., Kang,S., Bell,S.P. and MacAlpine,D.M. (2010) Conserved nucleosome positioning defines replication origins. *Genes Dev.*, **24**, 748–753.

153. Yadav,T. and Whitehouse,I. (2016) Replication-Coupled Nucleosome Assembly and Positioning by ATP-Dependent Chromatin-Remodeling Enzymes. *Cell Rep.*, **15**, 715–723.

154. Shim,E.Y., Hong,S.J., Oum,J.-H., Yanez,Y., Zhang,Y. and Lee,S.E. (2007) RSC Mobilizes Nucleosomes To Improve Accessibility of Repair Machinery to the Damaged Chromatin. *Mol. Cell. Biol.*, **27**, 1602–1613.

155. Rao,S.S.P., Huang,S.-C., Glenn St Hilaire,B., Engreitz,J.M., Perez,E.M., Kieffer-Kwon,K.-R., Sanborn,A.L., Johnstone,S.E., Bascom,G.D., Bochkov,I.D., *et al.* (2017) Cohesin Loss Eliminates All Loop Domains. *Cell*, **171**, 305-320.e24.

156. 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.

157. Dekker,J. and Mirny,L. (2016) The 3D Genome as Moderator of Chromosomal Communication. *Cell*, **164**, 1110–1121.

158. Andersson,R. and Sandelin,A. (2020) Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.*, **21**, 71–87.

159. Popay,T.M. and Dixon,J.R. (2022) Coming full circle: On the origin and evolution of the looping model for enhancer–promoter communication. *J. Biol. Chem.*, **298**, 102117.

160. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

161. Byrne,B.J., Davis,M.S., Yamaguchi,J., Bergsma,D.J. and Subramanian,K.N. (1983) Definition of the simian virus 40 early promoter region and demonstration of a host range bias in the enhancement effect of the simian virus 40 72-base-pair repeat. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 721–725.

162. Symmons,O., Pan,L., Remeseiro,S., Aktas,T., Klein,F., Huber,W. and Spitz,F. (2016) The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Dev. Cell*, **39**, 529–543.

163. van Steensel,B. and Furlong,E.E.M. (2019) The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.*, **20**, 327–337.

164. Kang,J., Kim,Y.W., Park,S., Kang,Y. and Kim,A. (2021) Multiple CTCF sites cooperate with each other to maintain a TAD for enhancer–promoter interaction in the β□globin locus. *FASEB J.*, **35**, e21768.

165. Ren,G., Jin,W., Cui,K., Rodrigez,J., Hu,G., Zhang,Z., Larson,D.R. and Zhao,K. (2017) CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Mol. Cell*, **67**, 1049-1058.e6.

166. Luo,H., Zhu,G., Eshelman,M.A., Fung,T.K., Lai,Q., Wang,F., Zeisig,B.B., Lesperance,J., Ma,X., Chen,S., *et al.* (2022) HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia. *Mol. Cell*, **82**, 833-851.e11.

167. Kubo,N., Ishii,H., Xiong,X., Bianco,S., Meitinger,F., Hu,R., Hocker,J.D., Conte,M., Gorkin,D., Yu,M., *et al.* (2021) Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.*, **28**, 152–161.

168. Islam,Z., Saravanan,B., Walavalkar,K., Farooq,U., Singh,A.K., Radhakrishnan,S., 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.

169. Luo,H., Yu,Q., Liu,Y., Tang,M., Liang,M., Zhang,D., Xiao,T.S., Wu,L., Tan,M., Ruan,Y., *et al.* (2020) LATS kinase–mediated CTCF phosphorylation and selective loss of genomic binding. *Sci. Adv.*, **6**, eaaw4651.

170. Doyle,B., Fudenberg,G., Imakaev,M. and Mirny,L.A. (2014) Chromatin Loops as Allosteric Modulators of Enhancer-Promoter Interactions. *PLoS Comput. Biol.*, **10**, e1003867.

171. Narendra,V., Rocha,P.P., An,D., Raviram,R., Skok,J.A., Mazzoni,E.O. and Reinberg,D. (2015) CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science.*, **347**, 1017–1021.

172. Lupiáñez,D.G., Kraft,K., Heinrich,V., Krawitz,P., Brancati,F., Klopocki,E., Horn,D., Kayserili,H., Opitz,J.M., Laxova,R., *et al.* (2015) Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, **161**, 1012–1025.

173. Willi,M., Yoo,K.H., Reinisch,F., Kuhns,T.M., Lee,H.K., Wang,C. and Hennighausen,L. (2017) Facultative CTCF sites moderate mammary super-enhancer activity and regulate juxtaposed gene in non-mammary cells. *Nat. Commun.*, **8**, 16069.

174. Lee,H.K., Willi,M., Wang,C., Yang,C.M., Smith,H.E., Liu,C. and Hennighausen,L. (2017) Functional assessment of CTCF sites at cytokine-sensing mammary enhancers using CRISPR/Cas9 gene editing in mice. *Nucleic Acids Res.*, **45**, 4606–4618.

175. Kabirova,E., Ryzhkova,A., Lukyanchikova,V., Khabarova,A., Korablev,A., Shnaider,T., Nuriddinov,M., Belokopytova,P., Smirnov,A., Khotskin,N. V., *et al.* (2024) TAD border deletion at the Kit locus causes tissue-specific ectopic activation of a neighboring gene. *Nat. Commun.*, **15**, 4521.

176. Kim,K.L., Rahme,G.J., Goel,V.Y., El Farran,C.A., Hansen,A.S. and Bernstein,B.E. (2024) Dissection of a CTCF topological boundary uncovers principles of enhancer-oncogene regulation. *Mol. Cell*, **84**, 1365-1376.e7.

177. Rodríguez-Carballo,E., Lopez-Delisle,L., Zhan,Y., Fabre,P.J., Beccari,L., El-Idrissi,I., Huynh,T.H.N., Ozadam,H., Dekker,J. and Duboule,D. (2017) The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.*, **31**, 2264–2281.

178. Beagan,J.A. and Phillips-Cremins,J.E. (2020) On the existence and functionality of topologically associating domains. *Nat. Genet.*, **52**, 8–16.

179. Hnisz,D., Weintraub,A.S., Day,D.S., Valton,A.-L., Bak,R.O., Li,C.H., Goldmann,J., Lajoie,B.R., Fan,Z.P., Sigova,A.A., *et al.* (2016) Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science.*, **351**, 1454–1458.

180. Flavahan,W.A., Drier,Y., Liau,B.B., Gillespie,S.M., Venteicher,A.S., Stemmer-Rachamimov,A.O., Suvà,M.L. and Bernstein,B.E. (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, **529**, 110–114.

181. Rajderkar,S., Barozzi,I., Zhu,Y., Hu,R., Zhang,Y., Li,B., Alcaina Caro,A., Fukuda-Yuzawa,Y., Kelman,G., Akeza,A., *et al.* (2023) Topologically associating domain boundaries are required for normal genome function. *Commun. Biol.*, **6**, 435.

182. Gruber,S., Haering,C.H. and Nasmyth,K. (2003) Chromosomal Cohesin Forms a Ring. *Cell*, **112**, 765–777.

183. Shi,Z., Gao,H., Bai,X. and Yu,H. (2020) Cryo-EM structure of the human cohesin-NIPBL-DNA complex. *Science.*, **368**, 1454–1459.

184. Zhang,N., Jiang,Y., Mao,Q., Demeler,B., Tao,Y.J. and Pati,D. (2013) Characterization of the Interaction between the Cohesin Subunits Rad21 and SA1/2. *PLoS One*, **8**, e69458.

185. Canudas,S. and Smith,S. (2009) Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *J. Cell Biol.*, **187**, 165–173.

186. 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.e4.

187. 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.

188. Petela,N.J., Gligoris,T.G., Metson,J., Lee,B.-G., Voulgaris,M., Hu,B., Kikuchi,S., Chapard,C., Chen,W., Rajendra,E., *et al.* (2018) Scc2 Is a Potent Activator of Cohesin's ATPase that Promotes Loading by Binding Scc1 without Pds5. *Mol. Cell*, **70**, 1134-1148.e7.

189. Hu,B., Itoh,T., Mishra,A., Katoh,Y., Chan,K.-L., Upcher,W., Godlee,C., Roig,M.B., Shirahige,K. and Nasmyth,K. (2011) ATP Hydrolysis Is Required for Relocating Cohesin from Sites Occupied by Its Scc2/4 Loading Complex. *Curr. Biol.*, **21**, 12–24.

190. Ciosk,R., Shirayama,M., Shevchenko,A., Tanaka,T., Toth,A., Shevchenko,A. and Nasmyth,K. (2000) Cohesin's Binding to Chromosomes Depends on a Separate Complex Consisting of Scc2 and Scc4 Proteins. *Mol. Cell*, **5**, 243–254.

191. Murayama,Y. and Uhlmann,F. (2014) Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature*, **505**, 367–371.

192. Gruber,S., Arumugam,P., Katou,Y., Kuglitsch,D., Helmhart,W., Shirahige,K. and Nasmyth,K. (2006) Evidence that Loading of Cohesin Onto Chromosomes Involves Opening of Its SMC Hinge. *Cell*, **127**, 523–537.

193. Murayama,Y. and Uhlmann,F. (2015) DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell*, **163**, 1628–1640.

194. Higashi,T.L., Eickhoff,P., Sousa,J.S., Locke,J., Nans,A., Flynn,H.R., Snijders,A.P., Papageorgiou,G., O'Reilly,N., Chen,Z.A., *et al.* (2020) A Structure-Based Mechanism for DNA Entry into the Cohesin Ring. *Mol. Cell*, **79**, 917-933.e9.

195. Abbondanzieri,E.A., Greenleaf,W.J., Shaevitz,J.W., Landick,R. and Block,S.M. (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature*, **438**, 460–465.

196. Qi,Z., Pugh,R.A., Spies,M. and Chemla,Y.R. (2013) Sequence-dependent base pair stepping dynamics in XPD helicase unwinding. *Elife*, **2**, e00334.

197. Stigler,J., Çamdere,G.Ö., Koshland,D.E. and Greene,E.C. (2016) Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Rep.*, **15**, 988–998.

198. Pradhan,B., Barth,R., Kim,E., Davidson,I.F., Bauer,B., van Laar,T., Yang,W., Ryu,J.-K., van der Torre,J., Peters,J.-M., *et al.* (2022) SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Rep.*, **41**, 111491.

199. Shintomi,K. and Hirano,T. (2009) Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl–Pds5 and Sgo1. *Genes Dev.*, **23**, 2224–2236.

200. Haarhuis,J.H.I., van der Weide,R.H., Blomen,V.A., Yáñez-Cuna,J.O., Amendola,M., van Ruiten,M.S., Krijger,P.H.L., Teunissen,H., Medema,R.H., van Steensel,B., *et al.* (2017) The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*, **169**, 693-707.e14.

201. 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.e8.

202. Bell,A.C. and Felsenfeld,G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482–485.

203. 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.

204. Nakahashi,H., Kwon,K.-R.K., Resch,W., Vian,L., Dose,M., Stavreva,D., Hakim,O., Pruett,N., Nelson,S., Yamane,A., *et al.* (2013) A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Rep.*, **3**, 1678–1689.

205. Kim,T.H., Abdullaev,Z.K., Smith,A.D., Ching,K.A., Loukinov,D.I., Green,R.D., Zhang,M.Q., Lobanenkov,V. V. and Ren,B. (2007) Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell*, **128**, 1231–1245.

206. 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.

207. Krishna,S.S., Majumdar,I. and Grishin,N. V (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.*, **31**, 532–550.

208. Wolfe,S.A., Nekludova,L. and Pabo,C.O. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Biochim Biophys Acta Gene Regul Mech.*, **29**, 183–212.

209. Fedotova,A.A., Bonchuk,A.N., Mogila,V.A. and Georgiev,P.G. (2017) C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors. *Acta Naturae*, **9**, 47–58.

210. Crewe,M., Segev,A., Rueda,R. and Madabhushi,R. (2024) Atypical Modes of CTCF Binding Facilitate Tissue-Specific and Neuronal Activity-Dependent Gene Expression States. *Mol. Neurobiol.*, **61**, 3240–3257.

211. 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.

212. Li,Y., Haarhuis,J.H.I., Sedeño Cacciatore,Á., Oldenkamp,R., van Ruiten,M.S., Willems,L., Teunissen,H., Muir,K.W., de Wit,E., Rowland,B.D., *et al.* (2020) The structural basis for cohesin–CTCF-anchored loops. *Nature*, **578**, 472–476.

213. Nora,E.P., Caccianini,L., Fudenberg,G., So,K., Kameswaran,V., Nagle,A., Uebersohn,A., Hajj,B., Saux,A. Le, Coulon,A., *et al.* (2020) Molecular basis of CTCF binding polarity in genome folding. *Nat. Commun.*, **11**, 5612.

214. Hara,K., Zheng,G., Qu,Q., Liu,H., Ouyang,Z., Chen,Z., Tomchick,D.R. and Yu,H. (2014) Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. *Nat. Struct. Mol. Biol.*, **21**, 864–870.

215. Wutz,G., Ladurner,R., St Hilaire,B.G., Stocsits,R.R., Nagasaka,K., Pignard,B., Sanborn,A., Tang,W., Várnai,C., Ivanov,M.P., *et al.* (2020) ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. *Elife*, **9**, e.52091.

216. Porter,H., Li,Y., Neguembor,M.V., Beltran,M., Varsally,W., Martin,L., Cornejo,M.T., Pezić,D., Bhamra,A., Surinova,S., *et al.* (2023) Cohesin-independent STAG proteins interact with RNA and R-loops and promote complex loading. *Elife*, **12**, e79386.

217. 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.

218. Analikwu,B.T., Deshayes,A., Torre,J. van der, Guérin,T., Katan,A.J., Béneut,C., Barth,R., Phipps,J., Scolari,V., Veaute,X., *et al.* (2023) Telomere protein arrays stall DNA loop extrusion by condensin. *bioRxiv*.

219. Huis in 't Veld,P.J., Herzog,F., Ladurner,R., Davidson,I.F., Piric,S., Kreidl,E., Bhaskara,V., Aebersold,R. and Peters,J.-M. (2014) Characterization of a DNA exit gate in the human cohesin ring. *Science.*, **346**, 968–972.

220. Kim,E., Barth,R. and Dekker,C. (2023) Looping the Genome with SMC Complexes. *Annu. Rev. Biochem.*, **92**, 15–41.

221. 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.

222. 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.

223. 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.e7.

224. Wei,C., Jia,L., Huang,X., Tan,J., Wang,M., Niu,J., Hou,Y., Sun,J., Zeng,P., Wang,J., *et al.* (2022) CTCF organizes inter-A compartment interactions through RYBP-dependent phase separation. *Cell Res.*, **32**, 744–760.

225. Saldaña-Meyer,R., González-Buendía,E., Guerrero,G., Narendra,V., Bonasio,R., Recillas-Targa,F. and Reinberg,D. (2014) CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.*, **28**, 723–734.

226. Hansen,A.S., Hsieh,T.-H.S., Cattoglio,C., Pustova,I., Saldaña-Meyer,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.e13.

227. Kung,J.T., Kesner,B., An,J.Y., Ahn,J.Y., Cifuentes-Rojas,C., Colognori,D., Jeon,Y., Szanto,A., del Rosario,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.

228. Pan,H., Jin,M., Ghadiyaram,A., Kaur,P., Miller,H.E., Ta,H.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.

229. Huber,J., Tanasie,N.-L., Zernia,S. and Stigler,J. (2024) Single-molecule imaging reveals a direct role of CTCF's zinc fingers in SA interaction and cluster-dependent RNA recruitment. *Nucleic Acids Res.*, **52**, 6490–6506.

230. Gutierrez-Escribano,P., Newton,M.D., Llauró,A., Huber,J., Tanasie,L., Davy,J., Aly,I., Aramayo,R., Montoya,A., Kramer,H., *et al.* (2019) A conserved ATP- and Scc2/4-dependent activity for cohesin in tethering DNA molecules. *Sci. Adv.*, **5**, eaay6804.

231. Horsfield,J.A. (2023) Full circle: a brief history of cohesin and the regulation of gene expression. *FEBS J.*, **290**, 1670–1687.

232. Litwin,I., Pilarczyk,E. and Wysocki,R. (2018) The Emerging Role of Cohesin in the DNA Damage Response. *Genes (Basel).*, **9**, 581.

233. Nasmyth,K. (2001) Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis. *Annu. Rev. Genet.*, **35**, 673–745.

234. Bastié,N., Chapard,C., Dauban,L., Gadal,O., Beckouët,F. and Koszul,R. (2022) Smc3 acetylation, Pds5 and Scc2 control the translocase activity that establishes cohesin-dependent chromatin loops. *Nat. Struct. Mol. Biol.*, **29**, 575–585.

235. Hauf,S., Waizenegger,I.C. and Peters,J.-M. (2001) Cohesin Cleavage by Separase Required for Anaphase and Cytokinesis in Human Cells. *Science.*, **293**, 1320–1323.

236. Zhang,N., Kuznetsov,S.G., Sharan,S.K., Li,K., Rao,P.H. and Pati,D. (2008) A handcuff model for the cohesin complex. *J. Cell Biol.*, **183**, 1019–1031.

237. Murayama,Y., Samora,C.P., Kurokawa,Y., Iwasaki,H. and Uhlmann,F. (2018) Establishment of DNA-DNA Interactions by the Cohesin Ring. *Cell*, **172**, 465-477.e15.

238. Ünal,E., Arbel-Eden,A., Sattler,U., Shroff,R., Lichten,M., Haber,J.E. and Koshland,D. (2004) DNA Damage Response Pathway Uses Histone Modification to Assemble a Double-Strand Break-Specific Cohesin Domain. *Mol. Cell*, **16**, 991–1002.

239. Dion,V., Kalck,V., Seeber,A., Schleker,T. and Gasser,S.M. (2013) Cohesin and the nucleolus constrain the mobility of spontaneous repair foci. *EMBO Rep.*, **14**, 984–991.

240. Ayra-Plasencia,J., Symington,L. and Machín,F. (2024) Cohesin still drives homologous recombination repair of DNA double-strand breaks in late mitosis. *Elife*, 10.7554/eLife.92706.1.

241. Phipps,J., Toulouze,M., Ducrot,C., Costa,R., Brocas,C. and Dubrana,K. (2024) Cohesin complex oligomerization maintains end-tethering at DNA double-strand breaks. *Nat. Cell Biol.*, 10.1038/s41556-024-01552-2.

242. Wang,H., Maurano,M.T., Qu,H., Varley,K.E., Gertz,J., Pauli,F., Lee,K., Canfield,T., Weaver,M., Sandstrom,R., *et al.* (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.*, **22**, 1680–1688.

243. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.-M., *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.

244. Hashimoto,H., Wang,D., Horton,J.R., Zhang,X., Corces,V.G. and Cheng,X. (2017) Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol. Cell*, **66**, 711-720.e3.

245. Jang,H., Shin,W., Lee,J. and Do,J. (2017) CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes (Basel).*, **8**, 148.

246. Flavahan,W.A., Drier,Y., Johnstone,S.E., Hemming,M.L., Tarjan,D.R., Hegazi,E., Shareef,S.J., Javed,N.M., Raut,C.P., Eschle,B.K., *et al.* (2019) Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature*, **575**, 229–233.

247. 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.

248. Klenova,E.M., Chernukhin,I. V., El-Kady,A., Lee,R.E., Pugacheva,E.M., Loukinov,D.I., Goodwin,G.H., Delgado,D., Filippova,G.N., León,J., *et al.* (2001) Functional Phosphorylation Sites in the C-Terminal Region of the Multivalent Multifunctional Transcriptional Factor CTCF. *Mol. Cell. Biol.*, **21**, 2221–2234.

249. Farrar,D., Rai,S., Chernukhin,I., Jagodic,M., Ito,Y., Yammine,S., Ohlsson,R., Murrell,A. and Klenova,E. (2010) Mutational Analysis of the Poly(ADP-Ribosyl)ation Sites of the Transcription Factor CTCF Provides an Insight into the Mechanism of Its Regulation by Poly(ADP-Ribosyl)ation. *Mol. Cell. Biol.*, **30**, 1199–1216.

250. Pavlaki,I., Docquier,F., Chernukhin,I., Kita,G., Gretton,S., Clarkson,C.T., Teif,V.B. and Klenova,E. (2018) Poly(ADP-ribosyl)ation associated changes in CTCF-chromatin binding and gene expression in breast cells. *Biochim Biophys Acta Gene Regul Mech.*, **1861**, 718–730.

251. Docquier,F., Kita,G.-X., Farrar,D., Jat,P., O'Hare,M., Chernukhin,I., Gretton,S., Mandal,A., Alldridge,L. and Klenova,E. (2009) Decreased Poly(ADP-Ribosyl)ation of CTCF, a Transcription Factor, Is Associated with Breast Cancer Phenotype and Cell Proliferation. *Clin. Cancer Res.*, **15**, 5762–5771.

252. Joshi,K., Smith,J.A., Perusinghe,N. and Monoghan,P. (1986) Cell proliferation in the human mammary epithelium. Differential contribution by epithelial and myoepithelial cells. *Am. J. Pathol.*, **124**, 199–206.

253. 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.* (2019) RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol. Cell*, **76**, 412-422.e5.

254. Pugacheva,E.M., Kubo,N., Loukinov,D., Tajmul,M., Kang,S., Kovalchuk,A.L., Strunnikov,A. V., Zentner,G.E., Ren,B. and Lobanenkov,V. V. (2020) CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 2020–2031.

255. Wutz,A. and Gribnau,J. (2007) X inactivation Xplained. *Curr. Opin. Genet. Dev.*, **17**, 387–393.

256. Bacher,C.P., Guggiari,M., Brors,B., Augui,S., Clerc,P., Avner,P., Eils,R. and Heard,E. (2006) Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat. Cell Biol.*, **8**, 293–299.

257. Wutz,A., Rasmussen,T.P. and Jaenisch,R. (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.*, **30**, 167–174.

258. Engreitz,J.M., Pandya-Jones,A., McDonel,P., Shishkin,A., Sirokman,K., Surka,C., Kadri,S., Xing,J., Goren,A., Lander,E.S., *et al.* (2013) The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science.*, **341**, 1237973.

259. 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**, 1537–1551.

260. 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.e24.

261. 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.

262. Vega,H., Waisfisz,Q., Gordillo,M., Sakai,N., Yanagihara,I., Yamada,M., van Gosliga,D., Kayserili,H., Xu,C., Ozono,K., *et al.* (2005) Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. *Nat. Genet.*, **37**, 468–470.

263. Liu,J. and Krantz,I. (2009) Cornelia de Lange syndrome, cohesin, and beyond. *Clin. Genet.*, **76**, 303–314.

264. Weiss,F.D., Calderon,L., Wang,Y.-F., Georgieva,R., Guo,Y., Cvetesic,N., Kaur,M., Dharmalingam,G., Krantz,I.D., Lenhard,B., *et al.* (2021) Neuronal genes deregulated in Cornelia de Lange Syndrome respond to removal and re-expression of cohesin. *Nat. Commun.*, **12**, 2919.

265. Mills,J.A., Herrera,P.S., Kaur,M., Leo,L., McEldrew,D., Tintos-Hernandez,J.A., Rajagopalan,R., Gagne,A., Zhang,Z., Ortiz-Gonzalez,X.R., *et al.* (2018) NIPBL+/− haploinsufficiency reveals a constellation of transcriptome disruptions in the pluripotent and cardiac states. *Sci. Rep.*, **8**, 1056.

266. Nie,Z., Gao,W., Zhang,Y., Hou,Y., Liu,J., Li,Z., Xue,W., Ye,X. and Jin,A. (2019) STAG2 loss-of-function mutation induces PD-L1 expression in U2OS cells. *Ann. Transl. Med.*, **7**, 127.

267. Adane,B., Alexe,G., Seong,B.K.A., Lu,D., Hwang,E.E., Hnisz,D., Lareau,C.A., Ross,L., Lin,S., Dela Cruz,F.S., *et al.* (2021) STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma. *Cancer Cell*, **39**, 827-844.e10.

268. Franke,M., Ibrahim,D.M., Andrey,G., Schwarzer,W., Heinrich,V., Schöpflin,R., Kraft,K., Kempfer,R., Jerković,I., Chan,W.-L., *et al.* (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, **538**, 265–269.

269. Katainen,R., Dave,K., Pitkänen,E., Palin,K., Kivioja,T., Välimäki,N., Gylfe,A.E., Ristolainen,H., Hänninen,U.A., Cajuso,T., *et al.* (2015) CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.*, **47**, 818–821.

270. Chen,W., Zeng,Y.C., Achinger-Kawecka,J., Campbell,E., Jones,A.K., Stewart,A.G., Khoury,A. and Clark,S.J. (2024) Machine learning enables pan-cancer identification of mutational hotspots at persistent CTCF binding sites. *Nucleic Acids Res.*, **52**, 8086–8099.

271. Fudenberg,G. and Pollard,K.S. (2019) Chromatin features constrain structural variation across evolutionary timescales. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 2175–2180.

272. Braccioli,L. and de Wit,E. (2019) CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.*, **63**, 157–165.

273. Gregor,A., Oti,M., Kouwenhoven,E.N., Hoyer,J., Sticht,H., Ekici,A.B., Kjaergaard,S., Rauch,A., Stunnenberg,H.G., Uebe,S., *et al.* (2013) De Novo Mutations in the Genome Organizer CTCF Cause Intellectual Disability. *Am. J. Hum. Genet.*, **93**, 124–131.

274. Zhang,J., Hu,G., Lu,Y., Ren,H., Huang,Y., Wen,Y., Ji,B., Wang,D., Wang,H., Liu,H., *et al.* (2024) CTCF mutation at R567 causes developmental disorders via 3D genome rearrangement and abnormal neurodevelopment. *Nat. Commun.*, **15**, 5524.

275. Aulmann,S., Bläker,H., Penzel,R., Rieker,R.J., Otto,H.F. and Sinn,H.P. (2003) CTCF Gene Mutations in Invasive Ductal Breast Cancer. *Breast Cancer Res. Treat.*, **80**, 347–352.

276. Zighelboim,I., Mutch,D.G., Knapp,A., Ding,L., Xie,M., Cohn,D.E. and Goodfellow,P.J. (2014) High Frequency Strand Slippage Mutations in CTCF in MSI-Positive Endometrial Cancers. *Hum. Mutat.*, **35**, 63–65.

277. Hildebrand,E.M. and Dekker,J. (2020) Mechanisms and Functions of Chromosome

Compartmentalization. *Trends Biochem. Sci.*, **45**, 385–396.

278. Heitz,E. (1928) “Das” Heterochromatin der Moose. In *I. Jahrb. f. wiss. Bot.*Vol. 69, pp. 762–818.

279. Padeken,J., Methot,S.P. and Gasser,S.M. (2022) Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nat. Rev. Mol. Cell Biol.*, **23**, 623–640.

280. Becker,J.S., McCarthy,R.L., Sidoli,S., Donahue,G., Kaeding,K.E., He,Z., Lin,S., Garcia,B.A. and Zaret,K.S. (2017) Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol. Cell*, **68**, 1023-1037.e15.

281. Suka,N., Suka,Y., Carmen,A.A., Wu,J. and Grunstein,M. (2001) Highly Specific Antibodies Determine Histone Acetylation Site Usage in Yeast Heterochromatin and Euchromatin. *Mol. Cell*, **8**, 473–479.

282. Kerem,B.-S., Goitein,R., Diamond,G., Cedar,H. and Marcus,M. (1984) Mapping of DNAase I sensitive regions on mitotic chromosomes. *Cell*, **38**, 493–499.

283. Gilbert,N. and Allan,J. (2001) Distinctive higher-order chromatin structure at mammalian centromeres. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 11949–11954.

284. Lawrence,M., Daujat,S. and Schneider,R. (2016) Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.*, **32**, 42–56.

285. Casas-Delucchi,C.S., van Bemmel,J.G., Haase,S., Herce,H.D., Nowak,D., Meilinger,D., Stear,J.H., Leonhardt,H. and Cardoso,M.C. (2012) Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. *Nucleic Acids Res.*, **40**, 159–169.

286. Janssen,A., Colmenares,S.U. and Karpen,G.H. (2018) Heterochromatin: Guardian of the Genome. *Annu. Rev. Cell Dev. Biol.*, **34**, 265–288.

287. Shevelyov,Y.Y. (2023) Interactions of Chromatin with the Nuclear Lamina and Nuclear Pore Complexes. *Int. J. Mol. Sci.*, **24**, 15771.

288. Vertii,A., Ou,J., Yu,J., Yan,A., Pagès,H., Liu,H., Zhu,L.J. and Kaufman,P.D. (2019) Two contrasting classes of nucleolus-associated domains in mouse fibroblast heterochromatin. *Genome Res.*, **29**, 1235–1249.

289. Neusser,M., Schubel,V., Koch,A., Cremer,T. and Müller,S. (2007) Evolutionarily conserved, cell type and species-specific higher order chromatin arrangements in interphase nuclei of primates. *Chromosoma*, **116**, 307–320.

290. Saccone,S., Federico,C. and Bernardi,G. (2002) Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. *Gene*, **300**, 169–178.

291. Solovei,I., Thanisch,K. and Feodorova,Y. (2016) How to rule the nucleus: divide et impera. *Curr. Opin. Cell Biol.*, **40**, 47–59.

292. Goetze,S., Mateos-Langerak,J., Gierman,H.J., de Leeuw,W., Giromus,O., Indemans,M.H.G., Koster,J., Ondrej,V., Versteeg,R. and van Driel,R. (2007) The Three-Dimensional Structure of Human Interphase Chromosomes Is Related to the Transcriptome Map. *Mol. Cell. Biol.*, **27**, 4475–4487.

293. Carter,K.C., Bowman,D., Carrington,W., Fogarty,K., McNeil,J.A., Fay,F.S. and Lawrence,J.B. (1993) A Three-Dimensional View of Precursor Messenger RNA Metabolism Within the Mammalian Nucleus. *Science.*, **259**, 1330–1335.

294. Chen,Y.-J.C., Koutelou,E. and Dent,S.Y.R. (2022) Now open: Evolving insights to the roles of lysine acetylation in chromatin organization and function. *Mol. Cell*, **82**, 716–727.

295. Eberharter,A. and Becker,P.B. (2002) Histone acetylation: a switch between repressive and permissive chromatin. *EMBO Rep.*, **3**, 224–229.

296. Bruce,K., Myers,F.A., Mantouvalou,E., Lefevre,P., Greaves,I., Bonifer,C., Tremethick,D.J., Thorne,A.W. and Crane-Robinson,C. (2005) The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. *Nucleic Acids Res.*, **33**, 5633–5639.

297. Pokholok,D.K., Harbison,C.T., Levine,S., Cole,M., Hannett,N.M., Lee,T.I., Bell,G.W., Walker,K., Rolfe,P.A., Herbolsheimer,E., *et al.* (2005) Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. *Cell*, **122**, 517–527.

298. Jiang,C. and Pugh,B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, **10**, 161–172.

299. Tse,C., Sera,T., Wolffe,A.P. and Hansen,J.C. (1998) Disruption of Higher-Order Folding by Core Histone Acetylation Dramatically Enhances Transcription of Nucleosomal Arrays by RNA Polymerase III. *Mol. Cell. Biol.*, **18**, 4629–4638.

300. Wang,X. and Hayes,J.J. (2008) Acetylation Mimics within Individual Core Histone Tail Domains Indicate Distinct Roles in Regulating the Stability of Higher-Order Chromatin Structure. *Mol. Cell. Biol.*, **28**, 227–236.

301. Zeng,L., Zhang,Q., Li,S., Plotnikov,A.N., Walsh,M.J. and Zhou,M.-M. (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature*, **466**, 258–262.

302. Hassan,A.H., Neely,K.E. and Workman,J.L. (2001) Histone Acetyltransferase Complexes Stabilize SWI/SNF Binding to Promoter Nucleosomes. *Cell*, **104**, 817–827.

303. Nightingale,K.P., Wellinger,R.E., Sogo,J.M. and Becker,P.B. (1998) Histone acetylation facilitates RNA polymerase II transcription of the Drosophila hsp26 gene in chromatin. *EMBO J.*, **17**, 2865–2876.

304. Morrison,O. and Thakur,J. (2021) Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. *Int. J. Mol. Sci.*, **22**, 6922.

305. Visser,A.E., Jaunin,F., Fakan,S. and Aten,J.A. (2000) High resolution analysis of interphase chromosome domains. *J. Cell Sci.*, **113**, 2585–2593.

306. Branco,M.R. and Pombo,A. (2006) Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *PLoS Biol.*, **4**, e138.

307. Cremer,T. and Cremer,C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, **2**, 292–301.

308. Cremer,M., Küpper,K., Wagler,B., Wizelman,L., von Hase,J., Weiland,Y., Kreja,L., Diebold,J., Speicher,M.R. and Cremer,T. (2003) Inheritance of gene density–related higher order chromatin arrangements in normal and tumor cell nuclei. *J. Cell Biol.*, **162**, 809–820.

309. Bolzer,A., Kreth,G., Solovei,I., Koehler,D., Saracoglu,K., Fauth,C., Müller,S., Eils,R., Cremer,C., Speicher,M.R., *et al.* (2005) Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. *PLoS Biol.*, **3**, e157.

310. Boyle,S., Gilchrist,S., Bridger,J.M., Mahy,N.L., Ellis,J.A. and Bickmore,W.A. (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.*, **10**, 211–9.

311. Dong,Q., Li,N., Li,X., Yuan,Z., Xie,D., Wang,X., Li,J., Yu,Y., Wang,J., Ding,B., *et al.* (2018) Genome□wide Hi□C analysis reveals extensive hierarchical chromatin interactions in rice. *Plant J.*, **94**, 1141–1156.

312. Shin,Y., Chang,Y.-C., Lee,D.S.W., Berry,J., Sanders,D.W., Ronceray,P., Wingreen,N.S., Haataja,M. and Brangwynne,C.P. (2018) Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell*, **175**, 1481-1491.e13.

313. Hnisz,D., Shrinivas,K., Young,R.A., Chakraborty,A.K. and Sharp,P.A. (2017) A Phase Separation Model for Transcriptional Control. *Cell*, **169**, 13–23.

314. Schwarzer,W., Abdennur,N., Goloborodko,A., Pekowska,A., Fudenberg,G., Loe-Mie,Y., Fonseca,N.A., Huber,W., Haering,C.H., Mirny,L., *et al.* (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, **551**, 51–56.

315. Ganser,L.R., Niaki,A.G., Yuan,X., Huang,E., Deng,D., Djaja,N.A., Ge,Y., Craig,A., Langlois,O. and Myong,S. (2024) The roles of FUS-RNA binding domain and low complexity domain in RNA-

dependent phase separation. *Structure.*, **32**, 177-187.e5.

316. Boija,A., Klein,I.A., Sabari,B.R., Dall'Agnese,A., Coffey,E.L., Zamudio,A. V., Li,C.H., Shrinivas,K., Manteiga,J.C., Hannett,N.M., *et al.* (2018) Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*, **175**, 1842-1855.e16.

317. Hyman,A.A., Weber,C.A. and Jülicher,F. (2014) Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.*, **30**, 39–58.

318. Frank,L. and Rippe,K. (2020) Repetitive RNAs as Regulators of Chromatin-Associated Subcompartment Formation by Phase Separation. *J. Mol. Biol.*, **432**, 4270–4286.

319. Sabari,B.R., Dall'Agnese,A., Boija,A., Klein,I.A., Coffey,E.L., Shrinivas,K., Abraham,B.J., Hannett,N.M., Zamudio,A. V., Manteiga,J.C., *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science.*, **361**, 379.

320. Azzaz,A.M., Vitalini,M.W., Thomas,A.S., Price,J.P., Blacketer,M.J., Cryderman,D.E., Zirbel,L.N., Woodcock,C.L., Elcock,A.H., Wallrath,L.L., *et al.* (2014) Human Heterochromatin Protein 1α Promotes Nucleosome Associations That Drive Chromatin Condensation. *J. Biol. Chem.*, **289**, 6850–6861.

321. Li,C.H., Coffey,E.L., Dall'Agnese,A., Hannett,N.M., Tang,X., Henninger,J.E., Platt,J.M., Oksuz,O., Zamudio,A. V, Afeyan,L.K., *et al.* (2020) MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature*, **586**, 440–444.

322. Zhang,H., Qin,W., Romero,H., Leonhardt,H. and Cardoso,M.C. (2023) Heterochromatin organization and phase separation. *Nucleus*, **14**, 2159142.

323. Okita,A.K., Zafar,F., Su,J., Weerasekara,D., Kajitani,T., Takahashi,T.S., Kimura,H., Murakami,Y., Masukata,H. and Nakagawa,T. (2019) Heterochromatin suppresses gross chromosomal rearrangements at centromeres by repressing Tfs1/TFIIS-dependent transcription. *Commun. Biol.*, **2**, 17.

324. Fanti,L., Giovinazzo,G., Berloco,M. and Pimpinelli,S. (1998) The Heterochromatin Protein 1 Prevents Telomere Fusions in Drosophila. *Mol. Cell*, **2**, 527–538.

325. Gaither,T.L., Merrett,S.L., Pun,M.J. and Scott,K.C. (2014) Centromeric Barrier Disruption Leads to Mitotic Defects in Schizosaccharomyces pombe. *G3 (Bethesda).*, **4**, 633–642.

326. Burns,K.H. (2017) Transposable elements in cancer. *Nat. Rev. Cancer*, **17**, 415–424.

327. Chojnowski,A., Ong,P.F., Foo,M.X.R., Liebl,D., Hor,L.-P., Stewart,C.L. and Dreesen,O. (2020) Heterochromatin loss as a determinant of progerin□induced DNA damage in Hutchinson–Gilford Progeria. *Aging Cell*, **19**, e13108.

328. Kilic,S., Felekyan,S., Doroshenko,O., Boichenko,I., Dimura,M., Vardanyan,H., Bryan,L.C., Arya,G., Seidel,C.A.M. and Fierz,B. (2018) Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α. *Nat. Commun.*, **9**, 235.

329. Keenen,M.M., Brown,D., Brennan,L.D., Renger,R., Khoo,H., Carlson,C.R., Huang,B., Grill,S.W., Narlikar,G.J. and Redding,S. (2021) HP1 proteins compact DNA into mechanically and positionally stable phase separated domains. *Elife*, **10**, e64563.

330. Her,C., Phan,T.M., Jovic,N., Kapoor,U., Ackermann,B.E., Rizuan,A., Kim,Y.C., Mittal,J. and Debelouchina,G.T. (2022) Molecular interactions underlying the phase separation of HP1α: role of phosphorylation, ligand and nucleic acid binding. *Nucleic Acids Res.*, **50**, 12702–12722.

331. Plys,A.J., Davis,C.P., Kim,J., Rizki,G., Keenen,M.M., Marr,S.K. and Kingston,R.E. (2019) Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.*, **33**, 799–813.

332. Eeftens,J.M., Kapoor,M., Michieletto,D. and Brangwynne,C.P. (2021) Polycomb condensates can promote epigenetic marks but are not required for sustained chromatin compaction. *Nat. Commun.*, **12**, 5888.

333. Boyle,S., Flyamer,I.M., Williamson,I., Sengupta,D., Bickmore,W.A. and Illingworth,R.S. (2020) A central role for canonical PRC1 in shaping the 3D nuclear landscape. *Genes Dev.*, **34**, 931–949.

334. Georgel,P.T., Horowitz-Scherer,R.A., Adkins,N., Woodcock,C.L., Wade,P.A. and Hansen,J.C. (2003) Chromatin Compaction by Human MeCP2. *J. Biol. Chem.*, **278**, 32181–32188.

335. Wang,L., Gao,Y., Zheng,X., Liu,C., Dong,S., Li,R., Zhang,G., Wei,Y., Qu,H., Li,Y., *et al.* (2019) Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. *Mol. Cell*, **76**, 646-659.e6.

336. Hubstenberger,A., Courel,M., Bénard,M., Souquere,S., Ernoult-Lange,M., Chouaib,R., Yi,Z., Morlot,J.-B., Munier,A., Fradet,M., *et al.* (2017) P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol. Cell*, **68**, 144-157.e5.

337. Schwartz,J.C., Ebmeier,C.C., Podell,E.R., Heimiller,J., Taatjes,D.J. and Cech,T.R. (2012) FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev.*, **26**, 2690–2695.

338. Burke,K.A., Janke,A.M., Rhine,C.L. and Fawzi,N.L. (2015) Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell*, **60**, 231–241.

339. Lu,H., Yu,D., Hansen,A.S., Ganguly,S., Liu,R., Heckert,A., Darzacq,X. and Zhou,Q. (2018) Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature*, **558**, 318–323.

340. Boehning,M., Dugast-Darzacq,C., Rankovic,M., Hansen,A.S., Yu,T., Marie-Nelly,H., McSwiggen,D.T., Kokic,G., Dailey,G.M., Cramer,P., *et al.* (2018) RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.*, **25**, 833–840.

341. Velychko,T., Mohammad,E., Ferrer-Vicens,I., Parfentev,I., Werner,M., Studniarek,C., Schwalb,B., Urlaub,H., Murphy,S., Cramer,P., *et al.* (2024) CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release. *Mol. Cell*, **84**, 2287-2303.e10.

342. Ørom,U.A., Derrien,T., Beringer,M., Gumireddy,K., Gardini,A., Bussotti,G., Lai,F., Zytnicki,M., Notredame,C., Huang,Q., *et al.* (2010) Long Noncoding RNAs with Enhancer-like Function in Human Cells. *Cell*, **143**, 46–58.

343. Li,W., Notani,D., Ma,Q., Tanasa,B., Nunez,E., Chen,A.Y., Merkurjev,D., Zhang,J., Ohgi,K., Song,X., *et al.* (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, **498**, 516–520.

344. Kim,T.-K., Hemberg,M., Gray,J.M., Costa,A.M., Bear,D.M., Wu,J., Harmin,D.A., Laptewicz,M., Barbara-Haley,K., Kuersten,S., *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182–187.

345. Rahnamoun,H., Lee,J., Sun,Z., Lu,H., Ramsey,K.M., Komives,E.A. and Lauberth,S.M. (2018) RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nat. Struct. Mol. Biol.*, **25**, 687–697.

346. Hah,N., Benner,C., Chong,L.-W., Yu,R.T., Downes,M. and Evans,R.M. (2015) Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E297-302.

347. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*, **153**, 307–319.

348. Guo,Y.E., Manteiga,J.C., Henninger,J.E., Sabari,B.R., Dall'Agnese,A., Hannett,N.M., Spille,J.-H., Afeyan,L.K., Zamudio,A. V, Shrinivas,K., *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*, **572**, 543–548.

349. Du,M., Stitzinger,S.H., Spille,J.-H., Cho,W.-K., Lee,C., Hijaz,M., Quintana,A. and Cissé,I.I. (2024) Direct observation of a condensate effect on super-enhancer controlled gene bursting. *Cell*, **187**, 331-344.e17.

350. Zamudio,A. V, Dall'Agnese,A., Henninger,J.E., Manteiga,J.C., Afeyan,L.K., Hannett,N.M., Coffey,E.L., Li,C.H., Oksuz,O., Sabari,B.R., *et al.* (2019) Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol. Cell*, **76**, 753-766.e6.

351. Tang,S.C., Vijayakumar,U., Zhang,Y. and Fullwood,M.J. (2022) Super-Enhancers, Phase-

Separated Condensates, and 3D Genome Organization in Cancer. *Cancers (Basel).*, **14**, 2866.

352. 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.

353. Lee,R., Kang,M.-K., Kim,Y.-J., Yang,B., Shim,H., Kim,S., Kim,K., Yang,C.M., Min,B., Jung,W.-J., *et al.* (2022) CTCF-mediated chromatin looping provides a topological framework for the formation of phase-separated transcriptional condensates. *Nucleic Acids Res.*, **50**, 207–226.

354. Dowen,J.M., Fan,Z.P., Hnisz,D., Ren,G., Abraham,B.J., Zhang,L.N., Weintraub,A.S., Schuijers,J., Lee,T.I., Zhao,K., *et al.* (2014) Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell*, **159**, 374–387.

355. Khan,A., Mathelier,A. and Zhang,X. (2018) Super-enhancers are transcriptionally more active and cell type-specific than stretch enhancers. *Epigenetics*, **13**, 910–922.

356. Marbouty,M., Le Gall,A., Cattoni,D.I., Cournac,A., Koh,A., Fiche,J.-B., Mozziconacci,J., Murray,H., Koszul,R. and Nollmann,M. (2015) Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Mol. Cell*, **59**, 588–602.

357. Hsieh,T.-H.S., Weiner,A., Lajoie,B., Dekker,J., Friedman,N. and Rando,O.J. (2015) Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*, **162**, 108–119.

358. Le,T.B.K., Imakaev,M. V, Mirny,L.A. and Laub,M.T. (2013) High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science.*, **342**, 731–734.

359. Rowley,M.J. and Corces,V.G. (2016) The three-dimensional genome: principles and roles of long-distance interactions. *Curr. Opin. Cell Biol.*, **40**, 8–14.

360. Mizuguchi,T., Fudenberg,G., Mehta,S., Belton,J.-M., Taneja,N., Folco,H.D., FitzGerald,P., Dekker,J., Mirny,L., Barrowman,J., *et al.* (2014) Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature*, **516**, 432–435.

361. Crane,E., Bian,Q., McCord,R.P., Lajoie,B.R., Wheeler,B.S., Ralston,E.J., Uzawa,S., Dekker,J. and Meyer,B.J. (2015) Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*, **523**, 240–244.

362. Kaushal,A., Mohana,G., Dorier,J., Özdemir,I., Omer,A., Cousin,P., Semenova,A., Taschner,M., Dergai,O., Marzetta,F., *et al.* (2021) CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. *Nat. Commun.*, **12**, 1011.

363. Ulianov,S. V, Khrameeva,E.E., Gavrilov,A.A., Flyamer,I.M., Kos,P., Mikhaleva,E.A., Penin,A.A., Logacheva,M.D., Imakaev,M. V, Chertovich,A., *et al.* (2016) Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.*, **26**, 70–84.

364. Ramírez,F., Bhardwaj,V., Arrigoni,L., Lam,K.C., Grüning,B.A., Villaveces,J., Habermann,B., Akhtar,A. and Manke,T. (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.*, **9**, 189.

365. Lefevre,P., Witham,J., Lacroix,C.E., Cockerill,P.N. and Bonifer,C. (2008) The LPS-Induced Transcriptional Upregulation of the Chicken Lysozyme Locus Involves CTCF Eviction and Noncoding RNA Transcription. *Mol. Cell*, **32**, 129–139.

366. Banigan,E.J., Tang,W., van den Berg,A.A., Stocsits,R.R., Wutz,G., Brandão,H.B., Busslinger,G.A., Peters,J.-M. and Mirny,L.A. (2023) Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc. Natl. Acad. Sci. U.S.A.*, **120**, e2210480120.

367. Lengronne,A., Katou,Y., Mori,S., Yokobayashi,S., Kelly,G.P., Itoh,T., Watanabe,Y., Shirahige,K. and Uhlmann,F. (2004) Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature*, **430**, 573–578.

368. Busslinger,G.A., Stocsits,R.R., van der Lelij,P., Axelsson,E., Tedeschi,A., Galjart,N. and Peters,J.-M. (2017) Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*,

**544**, 503–507.

369. Glynn,E.F., Megee,P.C., Yu,H.-G., Mistrot,C., Unal,E., Koshland,D.E., DeRisi,J.L. and Gerton,J.L. (2004) Genome-Wide Mapping of the Cohesin Complex in the Yeast Saccharomyces cerevisiae. *PLoS Biol.*, **2**, e259.

370. 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.e22.

371. Kamakaka,R.T. and Thomas,J.O. (1990) Chromatin structure of transcriptionally competent and repressed genes. *EMBO J.*, **9**, 3997–4006.

372. Felsenfeld,G. (1978) Chromatin. *Nature*, **271**, 115–122.

373. Weintraub,H. and Groudine,M. (1976) Chromosomal Subunits in Active Genes Have an Altered Conformation. *Science.*, **193**, 848–856.

374. Kujirai,T. and Kurumizaka,H. (2020) Transcription through the nucleosome. *Curr. Opin. Struct. Biol.*, **61**, 42–49.

375. Holstege,F.C.P., Fiedler,U. and Timmers,H.T. (1997) Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.*, **16**, 7468–7480.

376. Zhan,Y., Grabbe,F., Oberbeckmann,E., Dienemann,C. and Cramer,P. (2024) Three-step mechanism of promoter escape by RNA polymerase II. *Mol. Cell*, **84**, 1699-1710.e6.

377. Luse,D.S. and Jacob,G.A. (1987) Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. *J. Biol. Chem.*, **262**, 14990–14997.

378. Hsu,L.M., Cobb,I.M., Ozmore,J.R., Khoo,M., Nahm,G., Xia,L., Bao,Y. and Ahn,C. (2006) Initial Transcribed Sequence Mutations Specifically Affect Promoter Escape Properties. *Biochemistry*, **45**, 8841–8854.

379. Tang,G.-Q., Roy,R., Bandwar,R.P., Ha,T. and Patel,S.S. (2009) Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 22175–22180.

380. Gong,P. and Martin,C.T. (2006) Mechanism of Instability in Abortive Cycling by T7 RNA Polymerase. *J. Biol. Chem.*, **281**, 23533–23544.

381. Revyakin,A., Liu,C., Ebright,R.H. and Strick,T.R. (2006) Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science.*, **314**, 1139–1143.

382. Koh,H.R., Roy,R., Sorokina,M., Tang,G.-Q., Nandakumar,D., Patel,S.S. and Ha,T. (2018) Correlating Transcription Initiation and Conformational Changes by a Single-Subunit RNA Polymerase with Near Base-Pair Resolution. *Mol. Cell*, **70**, 695-706.e5.

383. Laitem,C., Zaborowska,J., Tellier,M., Yamaguchi,Y., Cao,Q., Egloff,S., Handa,H. and Murphy,S. (2015) CTCF regulates NELF, DSIF and P-TEFb recruitment during transcription. *Transcription*, **6**, 79–90.

384. Diao,A.J., Su,B.G. and Vos,S.M. (2024) Pause Patrol: Negative Elongation Factor's Role in Promoter-Proximal Pausing and Beyond. *J. Mol. Biol.*, 10.1016/j.jmb.2024.168779.

385. Vos,S.M., Farnung,L., Urlaub,H. and Cramer,P. (2018) Structure of paused transcription complex Pol II–DSIF–NELF. *Nature*, **560**, 601–606.

386. Bernecky,C., Plitzko,J.M. and Cramer,P. (2017) Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nat. Struct. Mol. Biol.*, **24**, 809–815.

387. Marshall,N.F., Peng,J., Xie,Z. and Price,D.H. (1996) Control of RNA Polymerase II Elongation Potential by a Novel Carboxyl-terminal Domain Kinase. *J. Biol. Chem.*, **271**, 27176–27183.

388. Yamada,T., Yamaguchi,Y., Inukai,N., Okamoto,S., Mura,T. and Handa,H. (2006) P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. *Mol. Cell*, **21**, 227–237.

389. Fujinaga,K., Irwin,D., Huang,Y., Taube,R., Kurosu,T. and Peterlin,B.M. (2004) Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. *Mol. Cell. Biol.*, **24**, 787–795.

390. Gilchrist,D.A., Nechaev,S., Lee,C., Ghosh,S.K.B., Collins,J.B., Li,L., Gilmour,D.S. and Adelman,K. (2008) NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.*, **22**, 1921–1933.

391. Jimeno-González,S., Ceballos-Chávez,M. and Reyes,J.C. (2015) A positioned +1 nucleosome enhances promoter-proximal pausing. *Nucleic Acids Res.*, **43**, 3068–3078.

392. Kujirai,T., Ehara,H., Fujino,Y., Shirouzu,M., Sekine,S. and Kurumizaka,H. (2018) Structural basis of the nucleosome transition during RNA polymerase II passage. *Science.*, **362**, 595–598.

393. Bintu,L., Kopaczynska,M., Hodges,C., Lubkowska,L., Kashlev,M. and Bustamante,C. (2011) The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. *Nat. Struct. Mol. Biol.*, **18**, 1394–1399.

394. Studitsky,V.M., Kassavetis,G.A., Geiduschek,E.P. and Felsenfeld,G. (1997) Mechanism of Transcription Through the Nucleosome by Eukaryotic RNA Polymerase. *Science.*, **278**, 1960–1963.

395. Hodges,C., Bintu,L., Lubkowska,L., Kashlev,M. and Bustamante,C. (2009) Nucleosomal Fluctuations Govern the Transcription Dynamics of RNA Polymerase II. *Science.*, **325**, 626–628.

396. Filipovski,M., Soffers,J.H.M., Vos,S.M. and Farnung,L. (2022) Structural basis of nucleosome retention during transcription elongation. *Science.*, **376**, 1313–1316.

397. Studitsky,V.M., Clark,D.J. and Felsenfeld,G. (1994) A histone octamer can step around a transcribing polymerase without leaving the template. *Cell*, **76**, 371–382.

398. Kireeva,M.L., Walter,W., Tchernajenko,V., Bondarenko,V., Kashlev,M. and Studitsky,V.M. (2002) Nucleosome Remodeling Induced by RNA Polymerase II: Loss of the H2A/H2B Dimer during Transcription. *Mol. Cell*, **9**, 541–552.

399. Eriksson,P.R. and Clark,D.J. (2021) The yeast ISW1b ATP-dependent chromatin remodeler is critical for nucleosome spacing and dinucleosome resolution. *Sci. Rep.*, **11**, 4195.

400. Hsieh,F.-K., Kulaeva,O.I., Patel,S.S., Dyer,P.N., Luger,K., Reinberg,D. and Studitsky,V.M. (2013) Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 7654–7659.

401. Skene,P.J., Hernandez,A.E., Groudine,M. and Henikoff,S. (2014) The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *Elife*, **3**, e02042.

402. Epshtein,V., Toulmé,F., Rahmouni,A.R., Borukhov,S. and Nudler,E. (2003) Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J.*, **22**, 4719–4727.

403. Abdelkareem,M., Saint-André,C., Takacs,M., Papai,G., Crucifix,C., Guo,X., Ortiz,J. and Weixlbaumer,A. (2019) Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol. Cell*, **75**, 298-309.e4.

404. Hao,N., Donnelly,A.J., Dodd,I.B. and Shearwin,K.E. (2023) When push comes to shove - RNA polymerase and DNA-bound protein roadblocks. *Biophys. Rev.*, **15**, 355–366.

405. Qian,J., Cartee,A., Xu,W., Yan,Y., Wang,B., Artsimovitch,I., Dunlap,D. and Finzi,L. (2024) Reciprocating RNA Polymerase batters through roadblocks. *Nat. Commun.*, **15**, 3193.

406. Colin,J., Candelli,T., Porrua,O., Boulay,J., Zhu,C., Lacroute,F., Steinmetz,L.M. and Libri,D. (2014) Roadblock Termination by Reb1p Restricts Cryptic and Readthrough Transcription. *Mol. Cell*, **56**, 667–680.

407. Nudler,E. (2012) RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell*, **149**, 1438–1445.

408. Marr,M.T. and Roberts,J.W. (2000) Function of Transcription Cleavage Factors GreA and GreB at

a Regulatory Pause Site. *Mol. Cell*, **6**, 1275–1285.

409. Jin,J., Bai,L., Johnson,D.S., Fulbright,R.M., Kireeva,M.L., Kashlev,M. and Wang,M.D. (2010) Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. *Nat. Struct. Mol. Biol.*, **17**, 745–752.

410. Tei,S., Sakata,T., Yoshimura,A., Natsume,T., Kanemaki,M.T., Bando,M. and Shirahige,K. (2024) Cohesin regulates promoter-proximal pausing of RNA Polymerase II by limiting recruitment of super elongation complex. *bioRxiv*, 10.1101/2024.03.15.584908.

411. Wada,Y., Ohta,Y., Xu,M., Tsutsumi,S., Minami,T., Inoue,K., Komura,D., Kitakami,J., Oshida,N., Papantonis,A., *et al.* (2009) A wave of nascent transcription on activated human genes. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 18357–18361.

412. West,S., Gromak,N. and Proudfoot,N.J. (2004) Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, **432**, 522–525.

413. Girbig,M., Misiaszek,A.D. and Müller,C.W. (2022) Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nat. Rev. Mol. Cell Biol.*, **23**, 603–622.

414. Cortazar,M.A., Sheridan,R.M., Erickson,B., Fong,N., Glover-Cutter,K., Brannan,K. and Bentley,D.L. (2019) Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. *Mol. Cell*, **76**, 896-908.e4.

415. Kim,M., Krogan,N.J., Vasiljeva,L., Rando,O.J., Nedea,E., Greenblatt,J.F. and Buratowski,S. (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, **432**, 517–522.

416. Lopez Martinez,D. and Svejstrup,J.Q. (2024) Mechanisms of RNA Polymerase II Termination at the 3′-End of Genes. *J. Mol. Biol.*, 10.1016/j.jmb.2024.168735.

417. Hedouin,S., Logsdon,G.A., Underwood,J.G. and Biggins,S. (2022) A transcriptional roadblock protects yeast centromeres. *Nucleic Acids Res.*, **50**, 7801–7815.

418. Ling,Y.H. and Yuen,K.W.Y. (2019) Point centromere activity requires an optimal level of centromeric noncoding RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 6270–6279.

419. Hao,N., Krishna,S., Ahlgren-Berg,A., Cutts,E.E., Shearwin,K.E. and Dodd,I.B. (2014) Road rules for traffic on DNA—systematic analysis of transcriptional roadblocking in vivo. *Nucleic Acids Res.*, **42**, 8861–8872.

420. Yarrington,R.M., Richardson,S.M., Lisa Huang,C.R. and Boeke,J.D. (2012) Novel Transcript Truncating Function of Rap1p Revealed by Synthetic Codon-Optimized Ty1 Retrotransposon. *Genetics*, **190**, 523–535.

421. Tunitskaya,V.L. and Kochetkov,S.N. (2002) Review: Structural-functional analysis of bacteriophage T7 RNA polymerase. *Biochem. (Mosc).*, **67**, 1124–35.

422. Felsenfeld,G. and Groudine,M. (2003) Controlling the double helix. *Nature*, **421**, 448–453.

423. Clark,S.C., Chereji,R. V, Lee,P.R., Fields,R.D. and Clark,D.J. (2020) Differential nucleosome spacing in neurons and glia. *Neurosci. Lett.*, **714**, 134559.

424. Maeshima,K., Ide,S. and Babokhov,M. (2019) Dynamic chromatin organization without the 30-nm fiber. *Curr. Opin. Cell Biol.*, **58**, 95–104.

425. Kim,S. and Shendure,J. (2019) Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Mol. Cell*, **76**, 306–319.

426. O’Malley,R.C., Huang,S.C., Song,L., Lewsey,M.G., Bartlett,A., Nery,J.R., Galli,M., Gallavotti,A. and Ecker,J.R. (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*, **165**, 1280–1292.

427. Yin,Y., Morgunova,E., Jolma,A., Kaasinen,E., Sahu,B., Khund-Sayeed,S., Das,P.K., Kivioja,T., Dave,K., Zhong,F., *et al.* (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science.*, **356**, eaaj2239.

428. Zhu,F., Farnung,L., Kaasinen,E., Sahu,B., Yin,Y., Wei,B., Dodonova,S.O., Nitta,K.R.,

Morgunova,E., Taipale,M., *et al.* (2018) The interaction landscape between transcription factors and the nucleosome. *Nature*, **562**, 76–81.

429. Li,C., Rioux,K.L. and Delaney,S. (2022) Histone variants H3.3 and H2A.Z/H3.3 facilitate excision of uracil from nucleosome core particles. *DNA Repair (Amst).*, **116**, 103355.

430. Wang,H., Zhou,R. and Ji,X. (2023) Droplet formation assay for investigating phase-separation mechanisms of RNA Pol II transcription and CTCF functioning. *STAR Protoc.*, **4**, 102202.

431. Miller,H., Zhou,Z., Shepherd,J., Wollman,A.J.M. and Leake,M.C. (2018) Single-molecule techniques in biophysics: a review of the progress in methods and applications. *Reports Prog. Phys.*, **81**, 024601.

432. Walter,N.G. and Bustamante,C. (2014) Introduction to Single Molecule Imaging and Mechanics: Seeing and Touching Molecules One at a Time. *Chem. Rev.*, **114**, 3069–3071.

433. Cornish,P. V and Ha,T. (2007) A Survey of Single-Molecule Techniques in Chemical Biology. *ACS Chem. Biol.*, **2**, 53–61.

434. Freitag,M., Jaklin,S., Padovani,F., Radzichevici,E., Zernia,S., Schmoller,K.M. and Stigler,J. (2022) Single-molecule experiments reveal the elbow as an essential folding guide in SMC coiled-coil arms. *Biophys. J.*, **121**, 4702–4713.

435. Countryman,P., Fan,Y., Gorthi,A., Pan,H., Strickland,E., Kaur,P., Wang,X., Lin,J., Lei,X., White,C., *et al.* (2018) Cohesin SA2 is a sequence-independent DNA-binding protein that recognizes DNA replication and repair intermediates. *J. Biol. Chem.*, **293**, 1054–1069.

436. Young,G., Hundt,N., Cole,D., Fineberg,A., Andrecka,J., Tyler,A., Olerinyova,A., Ansari,A., Marklund,E.G., Collier,M.P., *et al.* (2018) Quantitative mass imaging of single biological macromolecules. *Science.*, **360**, 423–427.

437. 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.

438. Tam,J. and Merino,D. (2015) Stochastic optical reconstruction microscopy (STORM) in comparison with stimulated emission depletion (STED) and other imaging methods. *J. Neurochem.*, **135**, 643–658.

439. Carravilla,P., Dasgupta,A., Zhurgenbayeva,G., Danylchuk,D.I., Klymchenko,A.S., Sezgin,E. and Eggeling,C. (2021) Long-term STED imaging of membrane packing and dynamics by exchangeable polarity-sensitive dyes. *Biophys. Reports*, **1**, 100023.

440. Rimoli,C.V., Valades-Cruz,C.A., Curcio,V., Mavrakis,M. and Brasselet,S. (2022) 4polar-STORM polarized super-resolution imaging of actin filament organization in cells. *Nat. Commun.*, **13**, 301.

441. Asor,R. and Kukura,P. (2022) Characterising biomolecular interactions and dynamics with mass photometry. *Curr. Opin. Chem. Biol.*, **68**, 102132.

442. Sonn-Segev,A., Belacic,K., Bodrug,T., Young,G., VanderLinden,R.T., Schulman,B.A., Schimpf,J., Friedrich,T., Dip,P.V., Schwartz,T.U., *et al.* (2020) Quantifying the heterogeneity of macromolecular machines by mass photometry. *Nat. Commun.*, **11**, 1772.

443. Hundt,N., Cole,D., Hantke,M.F., Miller,J.J., Struwe,W.B. and Kukura,P. (2022) Direct observation of the molecular mechanism underlying protein polymerization. *Sci. Adv.*, **8**, eabm7935.

444. Bigelyte,G., Young,J.K., Karvelis,T., Budre,K., Zedaveinyte,R., Djukanovic,V., Van Ginkel,E., Paulraj,S., Gasior,S., Jones,S., *et al.* (2021) Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nat. Commun.*, **12**, 6191.

445. Liu,S. and Wang,Y. (2010) Application of AFM in microbiology: a review. *Scanning*, **32**, 61–73.

446. Nagao,E. and Dvorak,J.A. (1999) Phase Imaging by Atomic Force Microscopy: Analysis of Living Homoiothermic Vertebrate Cells. *Biophys. J.*, **76**, 3289–3297.

447. Dumitru,A.C. and Koehler,M. (2023) Recent advances in the application of atomic force microscopy to structural biology. *J. Struct. Biol.*, **215**, 107963.

448. Niina,T., Fuchigami,S. and Takada,S. (2020) Flexible Fitting of Biomolecular Structures to Atomic Force Microscopy Images via Biased Molecular Simulations. *J. Chem. Theory Comput.*, **16**, 1349–1358.

449. Pan,Y., Sun,Z., Maiti,A., Kanai,T., Matsuo,H., Li,M., Harris,R.S., Shlyakhtenko,L.S. and Lyubchenko,Y.L. (2017) Nanoscale Characterization of Interaction of APOBEC3G with RNA. *Biochemistry*, **56**, 1473–1481.

450. Kaur,P., Lu,X., Xu,Q., Irvin,E.M., Pappas,C., Zhang,H., Finkelstein,I.J., Shi,Z., Tao,Y.J., Yu,H., *et al.* (2023) High-speed AFM imaging reveals DNA capture and loop extrusion dynamics by cohesin-NIPBL. *J. Biol. Chem.*, **299**, 105296.

451. Bustamante,C.J., Chemla,Y.R., Liu,S. and Wang,M.D. (2021) Optical tweezers in single-molecule biophysics. *Nat. Rev. Methods Primers.*, **1**, 25.

452. Tanasie,N.-L., Gutiérrez-Escribano,P., Jaklin,S., Aragon,L. and Stigler,J. (2022) Stabilization of DNA fork junctions by Smc5/6 complexes revealed by single-molecule imaging. *Cell Rep.*, **41**, 111778.

453. Sun,M., Amiri,H., Tong,A.B., Shintomi,K., Hirano,T., Bustamante,C. and Heald,R. (2023) Monitoring the compaction of single DNA molecules in Xenopus egg extract in real time. *Proc. Natl. Acad. Sci. U.S.A.*, **120**, e2221309120.

454. Meinhold,S., Bauer,D., Huber,J., Merkel,U., Weißl,A., Žoldák,G. and Rief,M. (2019) An Active, Ligand-Responsive Pulling Geometry Reports on Internal Signaling between Subdomains of the DnaK Nucleotide-Binding Domain in Single-Molecule Mechanical Experiments. *Biochemistry*, **58**, 4744–4750.

455. Suksombat,S., Khafizov,R., Kozlov,A.G., Lohman,T.M. and Chemla,Y.R. (2015) Structural dynamics of E. coli single-stranded DNA binding protein reveal DNA wrapping and unwrapping pathways. *Elife*, **4**, e08193.

456. Comstock,M.J., Whitley,K.D., Jia,H., Sokoloski,J., Lohman,T.M., Ha,T. and Chemla,Y.R. (2015) Direct observation of structure-function relationship in a nucleic acid–processing enzyme. *Science.*, **348**, 352–354.

457. Whitley,K.D., Comstock,M.J. and Chemla,Y.R. (2017) High-Resolution Optical Tweezers Combined With Single-Molecule Confocal Microscopy. *Methods Enzymol.*, **582**, 137–169.

458. Axelrod,D. (2001) Total Internal Reflection Fluorescence Microscopy in Cell Biology. *Traffic*, **2**, 764–774.

459. Axelrod,D. (1981) Cell-substrate contacts illuminated by total internal reflection fluorescence. *J. Cell Biol.*, **89**, 141–145.

460. Zernia,S. and Stigler,J. (2024) DNA curtains for studying phase separation mechanisms of DNA-organizing proteins. *Methods Cell Biol.*, **182**, 95–108.

461. Casjens,S.R. and Hendrix,R.W. (2015) Bacteriophage lambda: Early pioneer and still relevant. *Virology*, **479**–**480**, 310–330.

462. Hershey,A.D., Burgi,E. and Ingraham,L. (1963) COHESION OF DNA MOLECULES ISOLATED FROM PHAGE LAMBDA. *Proc. Natl. Acad. Sci. U.S.A.*, **49**, 748–755.

463. Hohn,B., Wurtz,M., Klein,B., Lustig,A. and Hohn,T. (1974) Phage lambda DNA packaging, in vitro. *J. Supramol. Struct.*, **2**, 302–317.

464. Nozaki,S. (2022) Rapid and Accurate Assembly of Large DNA Assisted by In Vitro Packaging of Bacteriophage. *ACS Synth. Biol.*, **11**, 4113–4122.

465. Thomas,M., Cameron,J.R. and Davis,R.W. (1974) Viable Molecular Hybrids of Bacteriophage Lambda and Eukaryotic DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 4579–4583.

466. Hohn,B. (1979) In Vitro packaging of λ and cosmid DNA. In *Methods in enzymology*.Vol. 68, pp. 299–309.

467. Kang,Y., Cheon,N.Y., Cha,J., Kim,A., Kim,H., Lee,L., Kim,K.O., Jo,K. and Lee,J.Y. (2020) High-

throughput single-molecule imaging system using nanofabricated trenches and fluorescent DNA-binding proteins. *Biotechnol. Bioeng.*, **117**, 1640–1648.

468. Fazio,T.A., Lee,J.Y., Wind,S.J. and Greene,E.C. (2012) Assembly of DNA Curtains Using Hydrogen Silsesquioxane As a Barrier to Lipid Diffusion. *Anal. Chem.*, **84**, 7613–7617.

469. Kim,S., Kim,Y. and Lee,J.Y. (2024) Real-time single-molecule visualization using DNA curtains reveals the molecular mechanisms underlying DNA repair pathways. *DNA Repair (Amst).*, **133**, 103612.

470. Cremer,P.S. and Boxer,S.G. (1999) Formation and Spreading of Lipid Bilayers on Planar Glass Supports. *J. Phys. Chem. B*, **103**, 2554–2559.

471. Fazio,T., Visnapuu,M.-L., Wind,S. and Greene,E.C. (2008) DNA Curtains and Nanoscale Curtain Rods: High-Throughput Tools for Single Molecule Imaging. *Langmuir*, **24**, 10524–10531.

472. Heyes,C.D., Groll,J., Möller,M. and Nienhaus,G.U. (2007) Synthesis, patterning and applications of star-shaped poly(ethylene glycol) biofunctionalized surfaces. *Mol. BioSyst.*, **3**, 419–430.

473. Maloney,A., Herskowitz,L.J. and Koch,S.J. (2011) Effects of Surface Passivation on Gliding Motility Assays. *PLoS One*, **6**, e19522.

474. Di Fiori,N. and Meller,A. (2009) Automated System for Single Molecule Fluorescence Measurements of Surface-immobilized Biomolecules. *J. Vis. Exp.*, **33**, 1542.

475. Collins,B.E., Ye,L.F., Duzdevich,D. and Greene,E.C. (2014) DNA curtains: novel tools for imaging protein-nucleic acid interactions at the single-molecule level. *Methods Cell Biol*, **123**, 217–234.

476. Schakenraad,K., Biebricher,A.S., Sebregts,M., Ten Bensel,B., Peterman,E.J.G., Wuite,G.J.L., Heller,I., Storm,C. and van der Schoot,P. (2017) Hyperstretching DNA. *Nat. Commun.*, **8**, 2197.

477. Lee,J.Y., Finkelstein,I.J., Crozat,E., Sherratt,D.J. and Greene,E.C. (2012) Single-molecule imaging of DNA curtains reveals mechanisms of KOPS sequence targeting by the DNA translocase FtsK. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 6531–6536.

478. Redding,S., Sternberg,S.H., Marshall,M., Gibb,B., Bhat,P., Guegler,C.K., Wiedenheft,B., Doudna,J.A. and Greene,E.C. (2015) Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell*, **163**, 854–865.

479. Silverstein,T.D., Gibb,B. and Greene,E.C. (2014) Visualizing protein movement on DNA at the single-molecule level using DNA curtains. *DNA Repair (Amst).*, **20**, 94–109.

480. Kang,H.J., Cheon,N.Y., Park,H., Jeong,G.W., Ye,B.J., Yoo,E.J., Lee,J.H., Hur,J.-H., Lee,E.-A., Kim,H., *et al.* (2021) TonEBP recognizes R-loops and initiates m6A RNA methylation for R-loop resolution. *Nucleic Acids Res.*, **49**, 269–284.

481. Cheon,N.Y., Kim,H.-S., Yeo,J.-E., Schärer,O.D. and Lee,J.Y. (2019) Single-molecule visualization reveals the damage search mechanism for the human NER protein XPC-RAD23B. *Nucleic Acids Res.*, **47**, 8337–8347.

482. Myler,L.R., Gallardo,I.F., Soniat,M.M., Deshpande,R.A., Gonzalez,X.B., Kim,Y., Paull,T.T. and Finkelstein,I.J. (2017) Single-Molecule Imaging Reveals How Mre11-Rad50-Nbs1 Initiates DNA Break Repair. *Mol. Cell*, **67**, 891-898.e4.

483. Zhou,H., Song,Z., Zhong,S., Zuo,L., Qi,Z., Qu,L.-J. and Lai,L. (2019) Mechanism of DNA□Induced Phase Separation for Transcriptional Repressor VRN1. *Angew. Chemie Int. Ed.*, **58**, 4858–4862.

484. Zuo,L., Ding,J. and Qi,Z. (2022) Visualizing carboxyl-terminal domain of RNA polymerase II recruitment by FET fusion protein condensates with DNA curtains. *Biophys. reports*, **8**, 80–89.

485. Soniat,M.M., Myler,L.R. and Finkelstein,I.J. (2019) Assembling the Human Resectosome on DNA Curtains. *Methods Mol Biol.*, **1999**, 225–244.

486. Kim,S., Shin,W.H., Kang,Y., Kim,H. and Lee,J.Y. (2024) Direct visualization of replication and R-loop collision using single-molecule imaging. *Nucleic Acids Res.*, **52**, 259–273.

487. Lee,J.Y., Finkelstein,I.J., Arciszewska,L.K., Sherratt,D.J. and Greene,E.C. (2014) Single-molecule imaging of FtsK translocation reveals mechanistic features of protein-protein collisions on DNA.

*Mol. Cell*, **54**, 832–43.

488. Rosano,G.L. and Ceccarelli,E.A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.*, **5**, 172.

489. Sørensen,H.P. and Mortensen,K.K. (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J. Biotechnol.*, **115**, 113–128.

490. Shis,D.L. and Bennett,M.R. (2014) Synthetic biology: the many facets of T7 RNA polymerase. *Mol. Syst. Biol.*, **10**, 745.

491. Borkotoky,S. and Murali,A. (2018) The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *Int. J. Biol. Macromol.*, **118**, 49–56.

492. McAllister,W.T., Morris,C., Rosenberg,A.H. and Studier,F.W. (1981) Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *J. Mol. Biol.*, **153**, 527–544.

493. CHAMBERLIN,M., McGRATH,J. and WASKELL,L. (1970) New RNA Polymerase from Escherichia coli infected with Bacteriophage T7. *Nature*, **228**, 227–231.

494. McAllister,W.T. and Wu,H.L. (1978) Regulation of transcription of the late genes of bacteriophage T7. *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 804–808.

495. Jeruzalmi,D. and Steitz,T.A. (1998) Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J.*, **17**, 4101–4113.

496. Cheetham,G.M.T. and Steitz, and T.A. (1999) Structure of a Transcribing T7 RNA Polymerase Initiation Complex. *Science.*, **286**, 2305–2309.

497. Imburgio,D., Rong,M., Ma,K. and McAllister,W.T. (2000) Studies of Promoter Recognition and Start Site Selection by T7 RNA Polymerase Using a Comprehensive Collection of Promoter Variants. *Biochemistry*, **39**, 10419–10430.

498. Cheetham,G.M., Jeruzalmi,D. and Steitz,T.A. (1999) Structural basis for initiation of transcription from an RNA polymerase–promoter complex. *Nature*, **399**, 80–83.

499. Dousis,A., Ravichandran,K., Hobert,E.M., Moore,M.J. and Rabideau,A.E. (2023) An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. *Nat. Biotechnol.*, **41**, 560–568.

500. Temiakov,D., Mentesana,P.E., Ma,K., Mustaev,A., Borukhov,S. and McAllister,W.T. (2000) The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 14109–14114.

501. Conrad,T., Plumbom,I., Alcobendas,M., Vidal,R. and Sauer,S. (2020) Maximizing transcription of nucleic acids with efficient T7 promoters. *Commun. Biol.*, **3**, 439.

502. Yin,Y.W. and Steitz,T.A. (2002) Structural Basis for the Transition from Initiation to Elongation Transcription in T7 RNA Polymerase. *Science.*, **298**, 1387–1395.

503. Tahirov,T.H., Temiakov,D., Anikin,M., Patlan,V., McAllister,W.T., Vassylyev,D.G. and Yokoyama,S. (2002) Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature*, **420**, 43–50.

504. Ma,K., Temiakov,D., Jiang,M., Anikin,M. and McAllister,W.T. (2002) Major Conformational Changes Occur during the Transition from an Initiation Complex to an Elongation Complex by T7 RNA Polymerase. *J. Biol. Chem.*, **277**, 43206–43215.

505. Lyakhov,D.L., He,B., Zhang,X., Studier,F.W., Dunn,J.J. and McAllister,W.T. (1998) Pausing and termination by bacteriophage T7 RNA polymerase. *J. Mol. Biol.*, **280**, 201–213.

506. He,B., Kukarin,A., Temiakov,D., Chin-Bow,S.T., Lyakhov,D.L., Rong,M., Durbin,R.K. and McAllister,W.T. (1998) Characterization of an Unusual, Sequence-specific Termination Signal for T7 RNA Polymerase. *J. Biol. Chem.*, **273**, 18802–18811.

507. Macdonald,L.E., Durbin,R.K., Dunn,J.J. and McAllister,W.T. (1994) Characterization of Two Types of Termination Signal for Bacteriophage T7 RNA Polymerase. *J. Mol. Biol.*, **238**, 145–158.

508. Ma,K., Temiakov,D., Anikin,M. and McAllister,W.T. (2005) Probing conformational changes in T7 RNA polymerase during initiation and termination by using engineered disulfide linkages. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 17612–17617.

509. Mead,D.A., Szczesna-Skorupa,E. and Kemper,B. (1986) Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *"Protein Eng.*, **1**, 67–74.

510. Kwon,Y.-S. and Kang,C. (1999) Bipartite Modular Structure of Intrinsic, RNA Hairpin-independent Termination Signal for Phage RNA Polymerases. *J. Biol. Chem.*, **274**, 29149–29155.

511. Song,H. and Kang,C. (2001) Sequence□specific termination by T7 RNA polymerase requires formation of paused conformation prior to the point of RNA release. *Genes to Cells*, **6**, 291–301.

512. Mukherjee,S., Brieba,L.G. and Sousa,R. (2003) Discontinuous movement and conformational change during pausing and termination by T7 RNA polymerase. *EMBO J.*, **22**, 6483–6493.

513. Zhang,X. and Studier,F.W. (2004) Multiple Roles of T7 RNA Polymerase and T7 Lysozyme During Bacteriophage T7 Infection. *J. Mol. Biol.*, **340**, 707–730.

514. Epshtein,V. and Nudler,E. (2003) Cooperation Between RNA Polymerase Molecules in Transcription Elongation. *Science.*, **300**, 801–805.

515. Chen,C., Xing,D., Tan,L., Li,H., Zhou,G., Huang,L. and Xie,X.S. (2017) Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science.*, **356**, 189–194.

516. Hashimshony,T., Senderovich,N., Avital,G., Klochendler,A., de Leeuw,Y., Anavy,L., Gennert,D., Li,S., Livak,K.J., Rozenblatt-Rosen,O., *et al.* (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.*, **17**, 77.

517. Sari,Y., Sousa Rosa,S., Jeffries,J. and Marques,M.P.C. (2024) Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production. *Sci. Rep.*, **14**, 9655.

518. Köhrer,C., Mayer,C., Gröbner,P. and Piendl,W. (1996) Use of T7 RNA Polymerase in an Optimized Escherichia coli Coupled in vitro Transcription□Translation System. Application in regulatory studies and expression of long transcription units. *Eur. J. Biochem.*, **236**, 234–239.

519. Kartje,Z.J., Janis,H.I., Mukhopadhyay,S. and Gagnon,K.T. (2021) Revisiting T7 RNA polymerase transcription in vitro with the Broccoli RNA aptamer as a simplified real-time fluorescent reporter. *J. Biol. Chem.*, **296**, 100175.

520. Kim,J.H. and Larson,R.G. (2007) Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules. *Nucleic Acids Res.*, **35**, 3848–3858.

521. 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.

522. Chen,Z. and Zhang,Y. (2005) Dimethyl sulfoxide targets phage RNA polymerases to promote transcription. *Biochem. Biophys. Res. Commun.*, **333**, 664–670.

523. Lyon,S. and Gopalan,V. (2018) A T7 RNA Polymerase Mutant Enhances the Yield of 5′□Thienoguanosine□Initiated RNAs. *ChemBioChem*, **19**, 142–146.

524. Guo,Q. and Sousa,R. (2005) Weakening of the T7 Promoter-Polymerase Interaction Facilitates Promoter Release. *J. Biol. Chem.*, **280**, 14956–14961.

525. Tang,G.-Q., Bandwar,R.P. and Patel,S.S. (2005) Extended Upstream A-T Sequence Increases T7 Promoter Strength. *J. Biol. Chem.*, **280**, 40707–40713.

526. Holden,S.J., Uphoff,S. and Kapanidis,A.N. (2011) DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nat. Methods*, **8**, 279–280.

527. Oh,S., Russell,S. and Sastry,S. (2009) Markov Chain Monte Carlo Data Association for Multi-Target Tracking. *IEEE Trans. Automat. Contr.*, **54**, 481–497.

528. Martin,F.H. and Tinoco,I. (1980) DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are

exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res.*, **8**, 2295–2300.

529. Naville,M., Ghuillot-Gaudeffroy,A., Marchais,A. and Gautheret,D. (2011) ARNold: A web tool for the prediction of Rho-independent transcription terminators. *RNA Biol.*, **8**, 11–13.

530. Park,J.-S. and Roberts,J.W. (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 4870–4875.

531. Santangelo,T.J. and Roberts,J.W. (2004) Forward Translocation Is the Natural Pathway of RNA Release at an Intrinsic Terminator. *Mol. Cell*, **14**, 117–126.

532. Zhou,Y., Navaroli,D.M., Enuameh,M.S. and Martin,C.T. (2007) Dissociation of halted T7 RNA polymerase elongation complexes proceeds via a forward-translocation mechanism. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 10352–10357.

533. Du,L., Villarreal,S. and Forster,A.C. (2012) Multigene expression in vivo: Supremacy of large versus small terminators for T7 RNA polymerase. *Biotechnol. Bioeng.*, **109**, 1043–1050.

534. Jung,Y. and Lippard,S.J. (2003) Multiple States of Stalled T7 RNA Polymerase at DNA Lesions Generated by Platinum Anticancer Agents. *J. Biol. Chem.*, **278**, 52084–52092.

535. 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.

536. Martin,C.T., Muller,D.K. and Coleman,J.E. (1988) Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry*, **27**, 3966–3974.

537. Weaver,T.M., Hoitsma,N.M., Spencer,J.J., Gakhar,L., Schnicker,N.J. and Freudenthal,B.D. (2022) Structural basis for APE1 processing DNA damage in the nucleosome. *Nat. Commun.*, **13**, 5390.

538. Lorch,Y., LaPointe,J.W. and Kornberg,R.D. (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, **49**, 203–210.

539. Trotta,E. (2022) GC content strongly influences the role of poly(dA) in the intrinsic nucleosome positioning in Saccharomyces cerevisiae. *Yeast*, **39**, 262–271.

540. Liu,G., Xing,Y., Zhao,H., Cai,L. and Wang,J. (2018) The implication of DNA bending energy for nucleosome positioning and sliding. *Sci. Rep.*, **8**, 8853.

541. Morozov,A. V, Fortney,K., Gaykalova,D.A., Studitsky,V.M., Widom,J. and Siggia,E.D. (2009) Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res.*, **37**, 4707–4722.

542. Hayes,J.J. and Hansen,J.C. (2002) New insights into unwrapping DNA from the nucleosome from a single-molecule optical tweezers method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 1752–1754.

543. Davey,C.A., Sargent,D.F., Luger,K., Maeder,A.W. and Richmond,T.J. (2002) Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. *J. Mol. Biol.*, **319**, 1097–1113.

544. Oguey,C., Foloppe,N. and Hartmann,B. (2010) Understanding the Sequence-Dependence of DNA Groove Dimensions: Implications for DNA Interactions. *PLoS One*, **5**, e15931.

545. Heddi,B., Oguey,C., Lavelle,C., Foloppe,N. and Hartmann,B. (2010) Intrinsic flexibility of B-DNA: the experimental TRX scale. *Nucleic Acids Res.*, **38**, 1034–1047.

546. Goodsell,D.S. and Dickerson,R.E. (1994) Bending and curvature calculations in B-DNA. *Nucleic Acids Res.*, **22**, 5497–5503.

547. Vinogradov,A.E. (2003) DNA helix: the importance of being GC-rich. *Nucleic Acids Res.*, **31**, 1838–1844.

548. Bernstein,B.E., Liu,C.L., Humphrey,E.L., Perlstein,E.O. and Schreiber,S.L. (2004) Global nucleosome occupancy in yeast. *Genome Biol.*, **5**, R62.

549. Zhang,Y., Moqtaderi,Z., Rattner,B.P., Euskirchen,G., Snyder,M., Kadonaga,J.T., Liu,X.S. and Struhl,K. (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome

positions in vivo. *Nat. Struct. Mol. Biol.*, **16**, 847–852.

550. Lee,J.Y. and Greene,E.C. (2011) Assembly of Recombinant Nucleosomes on Nanofabricated DNA Curtains for Single-Molecule Imaging. *Methods Mol. Biol.*, **778**, 243–258.

551. Tillo,D. and Hughes,T.R. (2009) G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics*, **10**, 442.

552. Gilchrist,D.A., Dos Santos,G., Fargo,D.C., Xie,B., Gao,Y., Li,L. and Adelman,K. (2010) Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. *Cell*, **143**, 540–551.

553. Frey,A., Listovsky,T., Guilbaud,G., Sarkies,P. and Sale,J.E. (2014) Histone H3.3 Is Required to Maintain Replication Fork Progression after UV Damage. *Curr. Biol.*, **24**, 2195–2201.

554. Colino-Sanguino,Y., Clark,S.J. and Valdes-Mora,F. (2022) The H2A.Z-nucleosome code in mammals: emerging functions. *Trends Genet.*, **38**, 273–289.

555. Giaimo,B.D., Ferrante,F., Herchenröther,A., Hake,S.B. and Borggrefe,T. (2019) The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin*, **12**, 37.

556. Gévry,N., Chan,H.M., Laflamme,L., Livingston,D.M. and Gaudreau,L. (2007) p21 transcription is regulated by differential localization of histone H2A.Z. *Genes Dev.*, **21**, 1869–1881.

557. Valdés-Mora,F., Gould,C.M., Colino-Sanguino,Y., Qu,W., Song,J.Z., Taylor,K.M., Buske,F.A., Statham,A.L., Nair,S.S., Armstrong,N.J., *et al.* (2017) Acetylated histone variant H2A.Z is involved in the activation of neo-enhancers in prostate cancer. *Nat. Commun.*, **8**, 1346.

558. Adam,M., Robert,F., Larochelle,M. and Gaudreau,L. (2001) H2A.Z Is Required for Global Chromatin Integrity and for Recruitment of RNA Polymerase II under Specific Conditions. *Mol. Cell. Biol.*, **21**, 6270–6279.

559. Rudnizky,S., Bavly,A., Malik,O., Pnueli,L., Melamed,P. and Kaplan,A. (2016) H2A.Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. *Nat. Commun.*, **7**, 12958.

560. Rong,Y.S. (2008) Loss of the Histone Variant H2A.Z Restores Capping to Checkpoint-Defective Telomeres in Drosophila. *Genetics*, **180**, 1869–1875.

561. Barski,A., Cuddapah,S., Cui,K., Roh,T.-Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, **129**, 823–837.

562. Brahma,S., Udugama,M.I., Kim,J., Hada,A., Bhardwaj,S.K., Hailu,S.G., Lee,T. and Bartholomew,B. (2017) INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nat. Commun.*, **8**, 15616.

563. Tramantano,M., Sun,L., Au,C., Labuz,D., Liu,Z., Chou,M., Shen,C. and Luk,E. (2016) Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. *Elife*, **5**, eLife.14243.

564. Jeronimo,C., Watanabe,S., Kaplan,C.D., Peterson,C.L. and Robert,F. (2015) The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations. *Mol. Cell*, **58**, 1113–1123.

565. Papamichos-Chronakis,M., Watanabe,S., Rando,O.J. and Peterson,C.L. (2011) Global Regulation of H2A.Z Localization by the INO80 Chromatin-Remodeling Enzyme Is Essential for Genome Integrity. *Cell*, **144**, 200–213.

566. Watanabe,S. and Peterson,C.L. (2016) Response to Comment on “A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme”. *Science.*, **353**, 358–358.

567. Wang,F., Ranjan,A., Wei,D. and Wu,C. (2016) Comment on “A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme”. *Science.*, **353**, 358–358.

568. Ranjan,A., Nguyen,V.Q., Liu,S., Wisniewski,J., Kim,J.M., Tang,X., Mizuguchi,G., Elalaoui,E., Nickels,T.J., Jou,V., *et al.* (2020) Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction. *Elife*, **9**, e55667.

569. Hatta,M. and Cirillo,L.A. (2007) Chromatin Opening and Stable Perturbation of Core Histone:DNA Contacts by FoxO1. *J. Biol. Chem.*, **282**, 35583–35593.

570. Cirillo,L.A., Lin,F.R., Cuesta,I., Friedman,D., Jarnik,M. and Zaret,K.S. (2002) Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Mol. Cell*, **9**, 279–289.

571. Bird,A., Taggart,M., Frommer,M., Miller,O.J. and Macleod,D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, **40**, 91–99.

572. Campanero,M.R., Armstrong,M.I. and Flemington,E.K. (2000) CpG methylation as a mechanism for the regulation of E2F activity. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6481–6486.

573. Thurman,R.E., Rynes,E., Humbert,R., Vierstra,J., Maurano,M.T., Haugen,E., Sheffield,N.C., Stergachis,A.B., Wang,H., Vernot,B., *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75–82.

574. Neph,S., Vierstra,J., Stergachis,A.B., Reynolds,A.P., Haugen,E., Vernot,B., Thurman,R.E., John,S., Sandstrom,R., Johnson,A.K., *et al.* (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, **489**, 83–90.

575. Barisic,D., Stadler,M.B., Iurlaro,M. and Schübeler,D. (2019) Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature*, **569**, 136–140.

576. Ooi,L., Belyaev,N.D., Miyake,K., Wood,I.C. and Buckley,N.J. (2006) BRG1 chromatin remodeling activity is required for efficient chromatin binding by repressor element 1-silencing transcription factor (REST) and facilitates REST-mediated repression. *J. Biol. Chem.*, **281**, 38974–38980.

577. Mehta,G.D., Ball,D.A., Eriksson,P.R., Chereji,R. V, Clark,D.J., McNally,J.G. and Karpova,T.S. (2018) Single-Molecule Analysis Reveals Linked Cycles of RSC Chromatin Remodeling and Ace1p Transcription Factor Binding in Yeast. *Mol. Cell*, **72**, 875-887.e9.

578. Soufi,A., Garcia,M.F., Jaroszewicz,A., Osman,N., Pellegrini,M. and Zaret,K.S. (2015) Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell*, **161**, 555–568.

579. King,H.W. and Klose,R.J. (2017) The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife*, **6**, e22631.

580. Fu,Y., Sinha,M., Peterson,C.L. and Weng,Z. (2008) The Insulator Binding Protein CTCF Positions 20 Nucleosomes around Its Binding Sites across the Human Genome. *PLoS Genet.*, **4**, e1000138.

581. Dynamic Biosensors GmbH (A Bruker Company) (2024) https://www.dynamic-biosensors.com/switchsense/.

582. Dynamic Biosensors GmbH (A Bruker Company) (2019) switchSENSE® Dynamic Response Sensing https://www.youtube.com/watch?v=RxIonG2cRJU 0:52.

583. Li,W., Shang,L., Huang,K., Li,J., Wang,Z. and Yao,H. (2017) Identification of critical base pairs required for CTCF binding in motif M1 and M2. *Protein Cell*, **8**, 544–549.

584. Brangwynne,C.P., Tompa,P. and Pappu,R. V. (2015) Polymer physics of intracellular phase transitions. *Nat. Phys.*, **11**, 899–904.

585. Glisovic,T., Bachorik,J.L., Yong,J. and Dreyfuss,G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.*, **582**, 1977–1986.

586. Brangwynne,C.P., Eckmann,C.R., Courson,D.S., Rybarska,A., Hoege,C., Gharakhani,J., Jülicher,F. and Hyman,A.A. (2009) Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science.*, **324**, 1729–1732.

587. Sunde,M., Serpell,L.C., Bartlam,M., Fraser,P.E., Pepys,M.B. and Blake,C.C.. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.*, **273**, 729–739.

588. Bienz,M. (2020) Head-to-Tail Polymerization in the Assembly of Biomolecular Condensates. *Cell*, **182**, 799–811.

589. Farag,M., Borcherds,W.M., Bremer,A., Mittag,T. and Pappu,R. V. (2023) Phase separation of

protein mixtures is driven by the interplay of homotypic and heterotypic interactions. *Nat. Commun.*, **14**, 5527.

590. Zhang,W., Cheng,L., Li,K., Xie,L., Ji,J., Lei,X., Jiang,A., Chen,C., Li,H., Li,P., *et al.* (2024) Evolutional heterochromatin condensation delineates chromocenter formation and retrotransposon silencing in plants. *Nat. Plants*, **10**, 1215–1230.

591. Caliari,S.R. and Burdick,J.A. (2016) A practical guide to hydrogels for cell culture. *Nat. Methods*, **13**, 405–414.

592. Ho,T.-C., Chang,C.-C., Chan,H.-P., Chung,T.-W., Shu,C.-W., Chuang,K.-P., Duh,T.-H., Yang,M.-H. and Tyan,Y.-C. (2022) Hydrogels: Properties and Applications in Biomedicine. *Molecules*, **27**, 2902.

593. Nath,P.C., Debnath,S., Sridhar,K., Inbaraj,B.S., Nayak,P.K. and Sharma,M. (2022) A Comprehensive Review of Food Hydrogels: Principles, Formation Mechanisms, Microstructure, and Its Applications. *Gels*, **9**, 1.

594. Erdel,F. and Rippe,K. (2018) Formation of Chromatin Subcompartments by Phase Separation. *Biophys. J.*, **114**, 2262–2270.

595. Ji,X., Dadon,D.B., Powell,B.E., Fan,Z.P., Borges-Rivera,D., Shachar,S., Weintraub,A.S., Hnisz,D., Pegoraro,G., Lee,T.I., *et al.* (2016) 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell*, **18**, 262–275.

596. Wassermann,L.M., Scheckenbach,M., Baptist,A. V, Glembockyte,V. and Heuer-Jungemann,A. (2023) Full Site-Specific Addressability in DNA Origami-Templated Silica Nanostructures. *Adv. Mater.*, **35**, e2212024.

597. Holzmann,J., Politi,A.Z., Nagasaka,K., Hantsche-Grininger,M., Walther,N., Koch,B., Fuchs,J., Dürnberger,G., Tang,W., Ladurner,R., *et al.* (2019) Absolute quantification of cohesin, CTCF and their regulators in human cells. *Elife*, **8**, e46269.

598. Chen,Z., Huang,J. and Li,J. (2022) Protein purification, crystallization, and structure determination of human DEAD-box RNA helicase DDX21 in different unwinding states. *STAR Protoc.*, **3**, 101642.

599. Hansma,H.G., Bezanilla,M., Zenhausern,F., Adrian,M. and Sinsheimer,R.L. (1993) Atomic force microscopy of DNA in aqueous solutions. *Nucleic Acids Res.*, **21**, 505–512.

600. Erickson,H.P. (2009) Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proced. Online*, **11**, 32–51.

601. 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.

602. 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.

603. Alberti,S., Gladfelter,A. and Mittag,T. (2019) Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*, **176**, 419–434.

604. Patel,A., Lee,H.O., Jawerth,L., Maharana,S., Jahnel,M., Hein,M.Y., Stoynov,S., Mahamid,J., Saha,S., Franzmann,T.M., *et al.* (2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*, **162**, 1066–77.

605. 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.

606. Kim,Y.J., Cecchini,K.R. and Kim,T.H. (2011) Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 7391–7396.

607. 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.

608. Rice,A. and Wereszczynski,J. (2017) Probing the disparate effects of arginine and lysine residues on antimicrobial peptide/bilayer association. *Biochim. Biophys. Acta - Biomembr.*, **1859**, 1941–1950.

609. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–74.

610. Silva,G.M., Maribo-Mogensen,B., Liang,X. and Kontogeorgis,G.M. (2024) Improving the Born equation: Origin of the Born radius and introducing dielectric saturation effects. *Fluid Phase Equilib.*, **576**, 113955.

611. Azevedo,C. and Saiardi,A. (2016) Why always lysine? The ongoing tale of one of the most modified amino acids. *Adv. Biol. Regul.*, **60**, 144–150.

612. Definitions and nomenclature of nucleic acid structure parameters. (1989) *EMBO J.*, **8**, 1–4.

613. Brukner,I., Dlakic,M., Savic,A., Susic,S., Pongor,S. and Suck,D. (1993) Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements. *Nucleic Acids Res.*, **21**, 1025–1029.

614. Grzeskowiak,K., Goodsell,D.S., Kaczor-Grzeskowiak,M., Cascio,D. and Dickerson,R.E. (1993) Crystallographic analysis of C-C-A-A-G-C-T-T-G-G and its implications for bending in B-DNA. *Biochemistry*, **32**, 8923–8931.

615. Goodsell,D.S., Kopka,M.L., Cascio,D. and Dickerson,R.E. (1993) Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 2930–2934.

616. Nelson,H.C.M., Finch,J.T., Luisi,B.F. and Klug,A. (1987) The structure of an oligo(dA)oligo(dT) tract and its biological implications. *Nature*, **330**, 221–226.

617. DiGabriele,A.D., Sanderson,M.R. and Steitz,T.A. (1989) Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 1816–1820.

618. Goodsell,D.S., Kaczor-Grzeskowiak,M. and Dickerson,R.E. (1994) The Crystal Structure of C-C-A-T-T-A-A-T-G-G. *J. Mol. Biol.*, **239**, 79–96.

619. Fonfría-Subirós,E., Acosta-Reyes,F., Saperas,N., Pous,J., Subirana,J.A. and Campos,J.L. (2012) Crystal Structure of a Complex of DNA with One AT-Hook of HMGA1. *PLoS One*, **7**, e37120.

620. Huth,J.R., Bewley,C.A., Nissen,M.S., Evans,J.N., Reeves,R., Gronenborn,A.M. and Clore,G.M. (1997) The solution structure of an HMG-I(Y)–DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.*, **4**, 657–665.

621. Rodríguez,J., Mosquera,J., Couceiro,J.R., Vázquez,M.E. and Mascareñas,J.L. (2015) The AT-Hook motif as a versatile minor groove anchor for promoting DNA binding of transcription factor fragments. *Chem. Sci.*, **6**, 4767–4771.

622. Reeves,R. and Nissen,M.S. (1990) The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J. Biol. Chem.*, **265**, 8573–8582.

623. Bisht,K.K., Daniloski,Z. and Smith,S. (2013) SA1 binds directly to DNA via its unique AT-hook to promote sister chromatid cohesion at telomeres. *J. Cell Sci.*, **126**, 3493–3503.

624. Roig,M.B., Löwe,J., Chan,K.-L., Beckouët,F., Metson,J. and Nasmyth,K. (2014) Structure and function of cohesin's Scc3/SA regulatory subunit. *FEBS Lett.*, **588**, 3692–3702.

625. Li,Y., Muir,K.W., Bowler,M.W., Metz,J., Haering,C.H. and Panne,D. (2018) Structural basis for Scc3-dependent cohesin recruitment to chromatin. *Elife*, **7**, e38356.

626. Kikuchi,S., Borek,D.M., Otwinowski,Z., Tomchick,D.R. and Yu,H. (2016) Crystal structure of the cohesin loader Scc2 and insight into cohesinopathy. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 12444–

12449.

627. Neuwald,A.F. and Hirano,T. (2000) HEAT Repeats Associated with Condensins, Cohesins, and Other Complexes Involved in Chromosome-Related Functions. *Genome Res.*, **10**, 1445–1452.

628. Williams,D.R., Lee,K.-J., Shi,J., Chen,D.J. and Stewart,P.L. (2008) Cryo-EM Structure of the DNA-Dependent Protein Kinase Catalytic Subunit at Subnanometer Resolution Reveals α Helices and Insight into DNA Binding. *Structure.*, **16**, 468–477.

629. Rubinson,E.H. and Eichman,B.F. (2012) Nucleic acid recognition by tandem helical repeats. *Curr. Opin. Struct. Biol.*, **22**, 101–109.

630. Yoshimura,S.H. and Hirano,T. (2016) HEAT repeats – versatile arrays of amphiphilic helices working in crowded environments? *J. Cell Sci.*, **129**, 3963–3970.

631. Piazza,I., Rutkowska,A., Ori,A., Walczak,M., Metz,J., Pelechano,V., Beck,M. and Haering,C.H. (2014) Association of condensin with chromosomes depends on DNA binding by its HEAT-repeat subunits. *Nat. Struct. Mol. Biol.*, **21**, 560–568.

632. Kschonsak,M., Merkel,F., Bisht,S., Metz,J., Rybin,V., Hassler,M. and Haering,C.H. (2017) Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell*, **171**, 588-600.e24.

633. Blat,Y. and Kleckner,N. (1999) Cohesins Bind to Preferential Sites along Yeast Chromosome III, with Differential Regulation along Arms versus the Centric Region. *Cell*, **98**, 249–259.

634. Borrie,M.S., Campor,J.S., Joshi,H. and Gartenberg,M.R. (2017) Binding, sliding, and function of cohesin during transcriptional activation. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E1062–E1071.

635. Weber,S.A., Gerton,J.L., Polancic,J.E., DeRisi,J.L., Koshland,D. and Megee,P.C. (2004) The kinetochore is an enhancer of pericentric cohesin binding. *PLoS Biol.*, **2**, E260.

636. Baker,R.E. and Rogers,K. (2005) Genetic and Genomic Analysis of the AT-Rich Centromere DNA Element II of Saccharomyces cerevisiae. *Genetics*, **171**, 1463–1475.

637. Štros,M. (2010) HMGB proteins: Interactions with DNA and chromatin. *Biochim Biophys Acta* , **1799**, 101–113.

638. Malarkey,C.S. and Churchill,M.E.A. (2012) The high mobility group box: the ultimate utility player of a cell. *Trends Biochem. Sci.*, **37**, 553–562.

639. Gerlitz,G., Hock,R., Ueda,T. and Bustin,M. (2009) The dynamics of HMG protein–chromatin interactions in living cells. *Biochem. Cell Biol.*, **87**, 127–137.

640. Thomas,J.O. and Travers,A.A. (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem. Sci.*, **26**, 167–174.

641. Klass,J., Murphy,F. V, Fouts,S., Serenil,M., Changela,A., Siple,J. and Churchill,M.E.A. (2003) The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity. *Nucleic Acids Res.*, **31**, 2852–2864.

642. Phair,R.D., Scaffidi,P., Elbi,C., Vecerová,J., Dey,A., Ozato,K., Brown,D.T., Hager,G., Bustin,M. and Misteli,T. (2004) Global Nature of Dynamic Protein-Chromatin Interactions In Vivo: Three-Dimensional Genome Scanning and Dynamic Interaction Networks of Chromatin Proteins. *Mol. Cell. Biol.*, **24**, 6393–6402.

643. Reddy,M.C., Christensen,J. and Vasquez,K.M. (2005) Interplay between Human High Mobility Group Protein 1 and Replication Protein A on Psoralen-Cross-linked DNA. *Biochemistry*, **44**, 4188–4195.

644. Ström,L., Lindroos,H.B., Shirahige,K. and Sjögren,C. (2004) Postreplicative Recruitment of Cohesin to Double-Strand Breaks Is Required for DNA Repair. *Mol. Cell*, **16**, 1003–1015.

645. Ahmadi,A., Rosnes,I., Blicher,P., Diekmann,R., Schüttpelz,M., Glette,K., Tørresen,J., Bjørås,M., Dalhus,B. and Rowe,A.D. (2018) Breaking the speed limit with multimode fast scanning of DNA by Endonuclease V. *Nat. Commun.*, **9**, 5381.

646. Dunn,A.R., Kad,N.M., Nelson,S.R., Warshaw,D.M. and Wallace,S.S. (2011) Single Qdot-labeled

glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.*, **39**, 7487–7498.

647. Blainey,P.C., van Oijen,A.M., Banerjee,A., Verdine,G.L. and Xie,X.S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 5752–5757.

648. Nelson,S.R., Dunn,A.R., Kathe,S.D., Warshaw,D.M. and Wallace,S.S. (2014) Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E2091-9.

649. Dalhus,B., Arvai,A.S., Rosnes,I., Olsen,Ø.E., Backe,P.H., Alseth,I., Gao,H., Cao,W., Tainer,J.A. and Bjørås,M. (2009) Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.*, **16**, 138–143.

650. Komazin-Meredith,G., Mirchev,R., Golan,D.E., van Oijen,A.M. and Coen,D.M. (2008) Hopping of a processivity factor on DNA revealed by single-molecule assays of diffusion. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 10721–10726.

651. Komazin-Meredith,G., Santos,W.L., Filman,D.J., Hogle,J.M., Verdine,G.L. and Coen,D.M. (2008) The Positively Charged Surface of Herpes Simplex Virus UL42 Mediates DNA Binding. *J. Biol. Chem.*, **283**, 6154–6161.

652. Hapgood,J.P., Riedemann,J. and Scherer,S.D. (2001) Regulation of gene expression by GC-rich DNA cis-elements. *Cell Biol. Int.*, **25**, 17–31.

653. Izhar,L., Adamson,B., Ciccia,A., Lewis,J., Pontano-Vaites,L., Leng,Y., Liang,A.C., Westbrook,T.F., Harper,J.W. and Elledge,S.J. (2015) A Systematic Analysis of Factors Localized to Damaged Chromatin Reveals PARP-Dependent Recruitment of Transcription Factors. *Cell Rep.*, **11**, 1486–1500.

654. Swaminathan,K., Flynn,P., Reece,R.J. and Marmorstein,R. (1997) Crystal structure of a PUT3–DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn2Cys6 binuclear cluster. *Nat. Struct. Biol.*, **4**, 751–759.

655. Van Roey,P., Waddling,C.A., Fox,K.M., Belfort,M. and Derbyshire,V. (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J.*, **20**, 3631–3637.

656. Bryk,M., Quirk,S.M., Mueller,J.E., Loizos,N., Lawrence,C. and Belfort,M. (1993) The td intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J.*, **12**, 2141–2149.

657. Meseguer,A., Årman,F., Fornes,O., Molina-Fernández,R., Bonet,J., Fernandez-Fuentes,N. and Oliva,B. (2020) On the prediction of DNA-binding preferences of C2H2-ZF domains using structural models: application on human CTCF. *NAR Genom Bioinform.*, **2**, lqaa046.

658. Hunter,C.A. and Lu,X.-J. (1997) DNA base-stacking interactions: a comparison of theoretical calculations with oligonucleotide X-ray crystal structures. *J. Mol. Biol.*, **265**, 603–619.

659. Slutsky,M. and Mirny,L.A. (2004) Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential. *Biophys. J.*, **87**, 4021–4035.

660. Iwahara,J. and Levy,Y. (2013) Speed-stability paradox in DNA-scanning by zinc-finger proteins. *Transcription*, **4**, 58–61.

661. Suter,D.M. (2020) Transcription Factors and DNA Play Hide and Seek. *Trends Cell Biol.*, **30**, 491–500.

662. Schmidt,H.G., Sewitz,S., Andrews,S.S. and Lipkow,K. (2014) An Integrated Model of Transcription Factor Diffusion Shows the Importance of Intersegmental Transfer and Quaternary Protein Structure for Target Site Finding. *PLoS One*, **9**, e108575.

663. Berg,O.G., Winter,R.B. and Von Hippel,P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.

664. Chen,J., Zhang,Z., Li,L., Chen,B.-C., Revyakin,A., Hajj,B., Legant,W., Dahan,M., Lionnet,T.,

Betzig,E., *et al.* (2014) Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. *Cell*, **156**, 1274–1285.

665. Feng,X.A., Yamadi,M., Fu,Y., Ness,K.M., Liu,C., Ahmed,I., Bowman,G.D., Johnson,M.E., Ha,T. and Wu,C. (2024) GAGA zinc finger transcription factor searches chromatin by 1D-3D facilitated diffusion. *bioRxiv Prepr. Serv. Biol.*, 10.1101/2023.07.14.549009.

666. Kent,S., Brown,K., Yang,C.-H., Alsaihati,N., Tian,C., Wang,H. and Ren,X. (2020) Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging. *Cell Rep.*, **33**, 108248.

667. Sugo,N., Morimatsu,M., Arai,Y., Kousoku,Y., Ohkuni,A., Nomura,T., Yanagida,T. and Yamamoto,N. (2015) Single-Molecule Imaging Reveals Dynamics of CREB Transcription Factor Bound to Its Target Sequence. *Sci. Rep.*, **5**, 10662.

668. de Jonge,W.J., Patel,H.P., Meeussen,J.V.W. and Lenstra,T.L. (2022) Following the tracks: How transcription factor binding dynamics control transcription. *Biophys. J.*, **121**, 1583–1592.

669. Halford,S.E. and Marko,J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–52.

670. Hammar,P., Leroy,P., Mahmutovic,A., Marklund,E.G., Berg,O.G. and Elf,J. (2012) The lac Repressor Displays Facilitated Diffusion in Living Cells. *Science.*, **336**, 1595–1598.

671. Tafvizi,A., Huang,F., Fersht,A.R., Mirny,L.A. and van Oijen,A.M. (2011) A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 563–568.

672. Mahmutovic,A., Berg,O.G. and Elf,J. (2015) What matters for lac repressor search in vivo—sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic Acids Res.*, **43**, 3454–3464.

673. Wang,Y.M., Austin,R.H. and Cox,E.C. (2006) Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.*, **97**, 048302.

674. Zandarashvili,L., Vuzman,D., Esadze,A., Takayama,Y., Sahu,D., Levy,Y. and Iwahara,J. (2012) Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E1724-32.

675. Doucleff,M. and Clore,G.M. (2008) Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 13871–13876.

676. Lieberman,B.A. and Nordeen,S.K. (1997) DNA intersegment transfer, how steroid receptors search for a target site. *J. Biol. Chem.*, **272**, 1061–1068.

677. Wollman,A.J., Shashkova,S., Hedlund,E.G., Friemann,R., Hohmann,S. and Leake,M.C. (2017) Transcription factor clusters regulate genes in eukaryotic cells. *Elife*, **6**, e27451.

678. Iwahara,J., Zweckstetter,M. and Clore,G.M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15062–15067.

679. Itoh,Y., Murata,A., Takahashi,S. and Kamagata,K. (2018) Intrinsically disordered domain of tumor suppressor p53 facilitates target search by ultrafast transfer between different DNA strands. *Nucleic Acids Res.*, **46**, 7261–7269.

680. 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.

681. Meeussen,J.V.W., Pomp,W., Brouwer,I., de Jonge,W.J., Patel,H.P. and Lenstra,T.L. (2023) Transcription factor clusters enable target search but do not contribute to target gene activation. *Nucleic Acids Res.*, **51**, 5449–5468.

682. Chen,X., Tsai,M.-Y. and Wolynes,P.G. (2022) The Role of Charge Density Coupled DNA Bending in Transcription Factor Sequence Binding Specificity: A Generic Mechanism for Indirect Readout. *J. Am. Chem. Soc.*, **144**, 1835–1845.

683. Hegde,R.S., Grossman,S.R., Laimins,L.A. and Sigler,P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DMA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.

684. Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.

685. Zhang,Y., Najmi,S.M. and Schneider,D.A. (2017) Transcription factors that influence RNA polymerases I and II: To what extent is mechanism of action conserved? *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1860**, 246–255.

686. Tolić-Nørrelykke,S.F., Rasmussen,M.B., Pavone,F.S., Berg-Sørensen,K. and Oddershede,L.B. (2006) Stepwise Bending of DNA by a Single TATA-Box Binding Protein. *Biophys. J.*, **90**, 3694–3703.

687. Coulombe,B. and Burton,Z.F. (1999) DNA Bending and Wrapping around RNA Polymerase: a "Revolutionary" Model Describing Transcriptional Mechanisms. *Microbiol. Mol. Biol. Rev.*, **63**, 457–478.

688. Kalendar,R. (2022) A Guide to Using FASTPCR Software for PCR, In Silico PCR, and Oligonucleotide Analysis. *Methods Mol Biol.*, **2392**, 223–243.

689. Mandel,M. and Marmur,J. (1968) [109] Use of ultraviolet absorbance-temperature profile for determining the guanine plus cytosine content of DNA. In *Methods in enzymology*.Vol. 12, pp. 195–206.

690. Luo,Y., North,J.A., Rose,S.D. and Poirier,M.G. (2014) Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res.*, **42**, 3017–3027.

691. Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **73**, 804–808.

692. Isalan,M. (2011) Zinc-finger nucleases: how to play two good hands. *Nat. Methods*, **9**, 32–34.

693. Finkelstein,I.J. and Greene,E.C. (2013) Molecular Traffic Jams on DNA. *Annu. Rev. Biophys.*, **42**, 241–263.

694. Whinn,K.S., Xu,Z.-Q., Jergic,S., Sharma,N., Spenkelink,L.M., Dixon,N.E., van Oijen,A.M. and Ghodke,H. (2023) Single-molecule visualization of stalled replication-fork rescue by the Escherichia coli Rep helicase. *Nucleic Acids Res.*, **51**, 3307–3326.

695. Chen,H., Tian,Y., Shu,W., Bo,X. and Wang,S. (2012) Comprehensive Identification and Annotation of Cell Type-Specific and Ubiquitous CTCF-Binding Sites in the Human Genome. *PLoS One*, **7**, e41374.

696. Hedglin,M. and O'Brien,P.J. (2010) Hopping Enables a DNA Repair Glycosylase To Search Both Strands and Bypass a Bound Protein. *ACS Chem. Biol.*, **5**, 427–436.

697. Cui,T.J., Klein,M., Hegge,J.W., Chandradoss,S.D., van der Oost,J., Depken,M. and Joo,C. (2019) Argonaute bypasses cellular obstacles without hindrance during target search. *Nat. Commun.*, **10**, 4390.

698. Altan-Bonnet,G., Libchaber,A. and Krichevsky,O. (2003) Bubble Dynamics in Double-Stranded DNA. *Phys. Rev. Lett.*, **90**, 138101.

699. Ambjörnsson,T., Banik,S.K., Krichevsky,O. and Metzler,R. (2007) Breathing Dynamics in Heteropolymer DNA. *Biophys. J.*, **92**, 2674–2684.

700. Inman,R.B. (1966) A denaturation map of the λ phage DNA molecule determined by electron microscopy. *J. Mol. Biol.*, **18**, 464–476.

701. Gorman,J., Plys,A.J., Visnapuu,M.-L., Alani,E. and Greene,E.C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.*, **17**, 932–938.

702. Barth,R., Pradhan,B., Kim,E., Davidson,I.F., van der Torre,J., Peters,J.-M. and Dekker,C. (2023) Testing pseudotopological and nontopological models for SMC-driven DNA loop extrusion against roadblock-traversal experiments. *Sci. Rep.*, **13**, 8100.

703. Finkelstein,I.J., Visnapuu,M.-L. and Greene,E.C. (2010) Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature*, **468**, 983–987.

704. Farah,J.A. and Smith,G.R. (1997) The RecBCD enzyme initiation complex for DNA unwinding: enzyme positioning and DNA opening. *J. Mol. Biol.*, **272**, 699–715.

705. Amundsen,S.K., Taylor,A.F. and Smith,G.R. (2020) Chi hotspot control of RecBCD helicase-nuclease by long-range intramolecular signaling. *Sci. Rep.*, **10**, 19415.

706. Schauer,G.D., Spenkelink,L.M., Lewis,J.S., Yurieva,O., Mueller,S.H., van Oijen,A.M. and O'Donnell,M.E. (2020) Replisome bypass of a protein-based R-loop block by Pif1. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 30354–30361.

707. Sugimoto,K., Kohara,Y. and Okazaki,T. (1987) Relative roles of T7 RNA polymerase and gene 4 primase for the initiation of T7 phage DNA replication in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 3977–3981.

708. Schweke,H., Pacesa,M., Levin,T., Goverde,C.A., Kumar,P., Duhoo,Y., Dornfeld,L.J., Dubreuil,B., Georgeon,S., Ovchinnikov,S., *et al.* (2024) An atlas of protein homo-oligomerization across domains of life. *Cell*, **187**, 999-1010.e15.

709. Nishi,H., Hashimoto,K., Madej,T. and Panchenko,A.R. (2013) Evolutionary, Physicochemical, and Functional Mechanisms of Protein Homooligomerization. *Prog Mol Biol Transl Sci.*, **117**, 3–24.

710. Sachl,R., Cujova,S., Singh,V., Riegerova,P., Hof,M., Steringer,J.P. and Nickel,W. (2022) In-membrane protein oligomerization as a critical step for membrane pore formation. *Biophys. J.*, **121**, 467a.

711. Peng,J., Ding,J., Tan,C., Baggenstoss,B., Zhang,Z., Lapolla,S.M. and Lin,J. (2009) Oligomerization of membrane-bound Bcl-2 is involved in its pore formation induced by tBid. *Apoptosis*, **14**, 1145–1153.

712. Mulvihill,E., Sborgi,L., Mari,S.A., Pfreundschuh,M., Hiller,S. and Müller,D.J. (2018) Mechanism of membrane pore formation by human gasdermin-D. *EMBO J.*, **37**, e98321.

713. Mesa-Galloso,H., Delgado-Magnero,K.H., Cabezas,S., López-Castilla,A., Hernández-González,J.E., Pedrera,L., Alvarez,C., Peter Tieleman,D., García-Sáez,A.J., Lanio,M.E., *et al.* (2017) Disrupting a key hydrophobic pair in the oligomerization interface of the actinoporins impairs their pore-forming activity. *Protein Sci.*, **26**, 550–565.

714. Nguyen,K.D.Q., Vigers,M., Sefah,E., Seppälä,S., Hoover,J.P., Schonenbach,N.S., Mertz,B., O'Malley,M.A. and Han,S. (2021) Homo-oligomerization of the human adenosine A2A receptor is driven by the intrinsically disordered C-terminus. *Elife*, **10**, e66662.

715. Ren,J., Wang,J., Ren,Y., Zhang,Y., Wei,P., Wang,M., Zhang,Y., Li,M., Yuan,C., Gong,H., *et al.* (2024) Structural basis of DNA recognition by BEN domain proteins reveals a role for oligomerization in unmethylated DNA selection by BANP. *Nucleic Acids Res.*, **52**, 11349–11361.

716. Litberg,T.J., Docter,B., Hughes,M.P., Bourne,J. and Horowitz,S. (2020) DNA Facilitates Oligomerization and Prevents Aggregation via DNA Networks. *Biophys. J.*, **118**, 162–171.

717. De Leo,A., Deng,Z., Vladimirova,O., Chen,H.-S., Dheekollu,J., Calderon,A., Myers,K.A., Hayden,J., Keeney,F., Kaufer,B.B., *et al.* (2019) LANA oligomeric architecture is essential for KSHV nuclear body formation and viral genome maintenance during latency. *PLOS Pathog.*, **15**, e1007489.

718. Soufi,A., Sawasdichai,A., Shukla,A., Noy,P., Dafforn,T., Smith,C., Jayaraman,P.-S. and Gaston,K. (2010) DNA compaction by the higher-order assembly of PRH/Hex homeodomain protein oligomers. *Nucleic Acids Res.*, **38**, 7513–7525.

719. Kovács,Z.J., Harami,G.M., Pálinkás,J., Kuljanishvili,N., Hegedüs,J., Harami-Papp,H., Mahmudova,L., Khamisi,L., Szakács,G. and Kovács,M. (2024) DNA-dependent phase separation by human SSB2 (NABP1/OBFC2A) protein points to adaptations to eukaryotic genome repair processes. *Protein Sci.*, **33**, e4959.

720. Ashton,N.W., Bolderson,E., Cubeddu,L., O'Byrne,K.J. and Richard,D.J. (2013) Human single-stranded DNA binding proteins are essential for maintaining genomic stability. *BMC Mol. Biol.*, **14**,

9.

721. Hnisz,D., Abraham,B.J., Lee,T.I., Lau,A., Saint-André,V., Sigova,A.A., Hoke,H.A. and Young,R.A. (2013) Super-Enhancers in the Control of Cell Identity and Disease. *Cell*, **155**, 934–947.

722. 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.

723. Eser,U., Chandler-Brown,D., Ay,F., Straight,A.F., Duan,Z., Noble,W.S. and Skotheim,J.M. (2017) Form and function of topologically associating genomic domains in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E3061–E3070.

724. Peña-Hernández,R., Marques,M., Hilmi,K., Zhao,T., Saad,A., Alaoui-Jamali,M.A., del Rincon,S. V, 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**, E677-686.

725. Wang,F., Tang,Z., Shao,H., Guo,J., Tan,T., Dong,Y. and Lin,L. (2018) Long noncoding RNA HOTTIP cooperates with CCCTC-binding factor to coordinate HOXA gene expression. *Biochem. Biophys. Res. Commun.*, **500**, 852–859.

726. Zhang,M., Huang,H., Li,J. and Wu,Q. (2024) ZNF143 deletion alters enhancer/promoter looping and CTCF/cohesin geometry. *Cell Rep.*, **43**, 113663.

727. Sun,X., Zhang,J. and Cao,C. (2022) CTCF and Its Partners: Shaper of 3D Genome during Development. *Genes (Basel).*, **13**, 1383.

728. Ishihara,K., Oshimura,M. and Nakao,M. (2006) CTCF-Dependent Chromatin Insulator Is Linked to Epigenetic Remodeling. *Mol. Cell*, **23**, 733–742.

# V Acknowledgements:

First, I want to thank Professor Johannes Stigler for being a great supervisor and for great discussions, ideas and explanations. He introduced me to the fascinating technique of DNA curtains and was always there for me when I had questions and needed new inspiration for my projects.

I also want to thank Dr. Sarah Zernia, who put a huge effort into supporting me during the publication process of the CTCF paper as well as this manuscript, helping me a great deal with figure preparation and improving my scientific writing skills. She also was always there for me when I had questions or needed help. I want to thank Dr. Marvin Freitag who helped me a lot with the preparation of this manuscript and had many inspiring ideas and suggestions. I want to thank Loredana Tanasie for being a great lab partner in our shared cohesin, CTCF and polymerase projects. Garp Linder I want to thank for the helpful discussion regarding DNA curtains and both him and Franziska Kuhnert for the help with the nucleosome chapter. I want to thank Dieter Kamp for helping me with IT and software problems. I want to thank Sigrun Jaklin, Jeanny Probst and Joelle Deplazes-Lauber for all of the great help in the lab. I want to thank everyone from the Stigler crew for having a great time in the lab. We had a great working atmosphere with everyone supporting each other and having an open ear whenever problems occurred. I also want to thank everyone in the lab for the inspiring scientific discussions we had and for the great fun during lab retreats and social events like going hiking or shared dinners and cocktail evenings.

I want to thank Veronika Bobinger and Günther Woehlke for introducing me to the switchsense® technology, Bingzhi Wang for the great help with analysing photobleaching data, and Daniel Bollschweiler for help with mass photometry measurements. I also want to thank Michael Scheckenbach, Johann Bohlen and Patrick Schüler from the Tinnefeld lab for performing AFM and graphene energy transfer measurements on CTCF.

I want to thank all of my friends and my family, who helped to stay motivated during the PhD process, by having a great time together during hiking trips, shared holidays, barbeque evenings or playing board games. I especially want to thank my parents for all the support during the years of my biochemistry study, making it possible for me to go to university and even start this PhD.