Modeling, visualizing and quantifying chromatin loop extrusion dynamics in living human cells

Thomas Sabaté

Sorbonne Université

Ecole doctorale 515 - Complexité du vivant

Imaging and modeling Unit (Institut Pasteur, Paris)

Cell biology of RNA (Institut de génétique humaine, Montpellier)

Modeling, visualizing and quantifying chromatin loop extrusion dynamics in living human cells

Par Thomas Sabaté

Thèse de doctorat de Génétique et génomique

Dirigée par Christophe Zimmer et Édouard Bertrand

Présentée et soutenue publiquement le 12 septembre 2023

Devant un jury composé de :

Christophe ZIMMER, Directeur de recherche	Co-directeur de thèse
Edouard BERTRAND, Directeur de recherche	Co-directeur de thèse
Gaelle LEGUBE, Directrice de recherche	Présidente du jury
Wendy BICKMORE, Professeur	Rapporteuse
Antoine COULON, Chargé de recherche	Examinateur
Mario NICODEMI, Professeur	Examinateur

Acknowledgements

I would like to thank Wendy Bickmore and Evi Soutoglou for accepting to evaluate my work and Antoine Coulon, Gaelle Legube and Mario Nicodemi for accepting to be part of my jury.

I also thank Marcelo Nollmann, Thomas Gregor and Auguste Genovesio who followed and provided valuable feedback on my work as members of my PhD committee.

I would like to thank my PhD supervisors, Edouard and Christophe, for initially accepting my PhD proposal, letting me work independently, and for the continuous discussions throughout my PhD. My work would have lacked a major part without one of you.

I would like to thank all past and present members of both labs, it was very nice and pleasing to always have a good atmosphere in the labs. I would like to have a special thanks to the colleagues that helped me during my work: Marie-Cécile for the painful acquisition of some of the cell lines (and for enduring my long emails and my critiques against Geslab), Jyotsana for introducing me to polymer simulations, Jean-Yves for the TrackMate improvements and Benoît Lelandais for his help with image processing and mathematical modelling. I am grateful to Héloïse Chassé for her comments on my manuscript, her help and advice with western blots and the beer nights in Montpellier. I would also like to deeply thank Marie Lemesle who was extremely helpful to get through all the administrative difficulties.

I would like to thank people from the Montpellier Ressources Imagerie (MRI), and especially Virginie Georget, Orestis Faklaris and Julio Mateos-Langerak for their constant help with the microscopes.

I would not have been able to do this work without all the teachers I met during my education. Especially, I would like to deeply thank the teachers from prépa BCPST Fénelon: Mme. Van der Rest, M. Clatin, M. Kerner, Mme. Ginestet, M. Anselme and M. Aubert. You taught me how to properly think and gave me a solid basis of knowledge in mathematics and physics, which have been very useful during my education and my PhD. Also, I would like to

thank Paolo de Los Rios at Ecole Polytechnique Fédérale de Lausanne whose course on statistical physics of macromolecules was extremely valuable.

Hélène, j'aurais été très fier de présenter ma thèse devant toi et te montrer mon premier papier. Je regrette de n'avoir pas pu plus discuter avec toi de *Listeria* et de science en général, malgré l'incompréhension totale qui régnait dans la famille autour de nous lorsque nous abordions le sujet.

Je tiens à remercier mes parents, pour leur aide dans les moments difficiles. J'aurais aimé pouvoir passer plus de temps avec vous. Maman, j'ai beaucoup aimé les soirées au théâtre, au cinéma ou les balades du week-end que l'on a faites. Je me souviendrai aussi de la période de confinement avec bonheur. Papa, j'ai toujours été étonné de ta constance et persévérance (plutôt fructueuses) à suivre et comprendre mon sujet de recherche.

Enfin, je voudrais remercier Lucile. Tu as enduré mes nombreuses absences physiques et mentales, à un point que j'ai toujours du mal à comprendre. Nos retrouvailles à Paris, Montpellier ou ailleurs m'ont toujours fait énormément de bien. Tu as été très attentionnée et compréhensive, tu m'as fait sortir de ma bulle dans les périodes de travail intenses et tu as réussi à le faire à merveille, aux bons moments, tout en comprenant ma passion pour la recherche. Nous avons passé de très beaux moments pendant ces années, beaucoup d'autres arrivent encore.

Table of contents

Introduction

1. The human genome is non-randomly organized at multiple scales

1.1 The human genome is highly folded in the cell nucleus.

1.1.1 Packing a 2-meter-long genome in a 6 µm diameter nucleus

The haploid human genome is made of 3.2 billion base pairs which are distributed across 23 chromosomes. The whole diploid genome is stored in each nucleus of almost all human cell types. It is organized as a double helix of 0.34 $^{nm}/_{bp}$ x 6.4x10^{9} bp ≈ 2 meters long. However, the whole genome is stored in a nucleus of only 6 µm diameter. Consequently, the genome is highly folded in the nucleus. Despite this tight folding, nuclear processes such as DNA replication, DNA repair and gene expression need to access the genome efficiently. The genome was found to be organized at different scales rather than randomly folded in the nucleus, as described in later sections.

1.1.2 Nucleosomes wrap DNA around them

The first level of genome compaction is achieved by histone proteins (**Figure 1A**). A histone protein octamer (two H2A-H2B dimers and a H3-H4 tetramer) together with DNA forms the nucleosome[1]. These cylindrical structures (11 nm diameter and 4 nm in height) wrap 1.7 turns of DNA around them, which equals 146 bp of DNA[1]. This interaction is passively made since the negatively charged DNA is attracted to the positively charged histones (a free energy gain of about a few dozens of k_bT)[2,3]. This complex of DNA and histones form the chromatin fiber. Although the interaction between histones and DNA is a passive mechanism, nucleosomes are dynamic structures that assemble, unwrap, slide and exchange histone variants[4].

This dynamic repositioning of nucleosomes is an energy-driven mechanism which is dependent on adenosine triphosphate (ATP, one of the cellular sources of energy) and occurs on the whole genome. The distribution of nucleosomes along the chromatin is well controlled. Indeed, nucleosomes are barriers for RNA Polymerase II (RNA Pol II), which allows the expression of genes by transcription[5] and they restrict the binding of transcription factors to DNA[6]. As a consequence, close to transcription start sites, the region just upstream of the

gene promoter is depleted from nucleosomes and downstream nucleosomes occupy specific and defined locations relative to the transcription start site[7–9].

Also, the tails of the histone proteins forming the nucleosome are accessible to enzymes that can add or remove covalent chemical bonds from the protein (respectively called 'writers' or 'erasers'). Therefore, each histone tail can receive a wide diversity of covalent protein modifications (acetylations (ac), methylations (me), ubiquitinations, sumoylation, etc.). The combination of such histone modifications creates the histone code[10,11]. This histone code defines different states of chromatin, which are associated to functional regions (**Figure 1A**). For instance, H3K27ac together with H3K4me1 are markers for transcriptional enhancers[9], while H3K27me3 and H3K9me3 marks define a repressive transcriptional environment[12]. Using this knowledge, it is possible to define chromatin states based on Chromatin ImmunoPrecipitation followed by next-generation sequencing (ChIP-Seq) data for histone modifications[13]. These histone marks can be bound by 'reader' proteins that affect the transcriptional state of chromatin by inducing either a repressive or a permissive chromatin environment[14]. These dynamically reversible changes can affect the transcriptional state of the chromatin both at short (hours and duration of a cell cycle[15]) and long timescales (memory across cell divisions and across individuals[16,17]). Histone modifications, together with DNA chemical modifications (*e.g.* methylation) are part of a larger group of mechanisms leading to changes in phenotypes without changes in genotypes called epigenetics[18].

Thus, nucleosomes form the first level of genome compaction, but their role is not solely structural. Indeed, based on their genome-wide occupancy and the chemical histone modifications, nucleosomes define functional regions of chromatin.

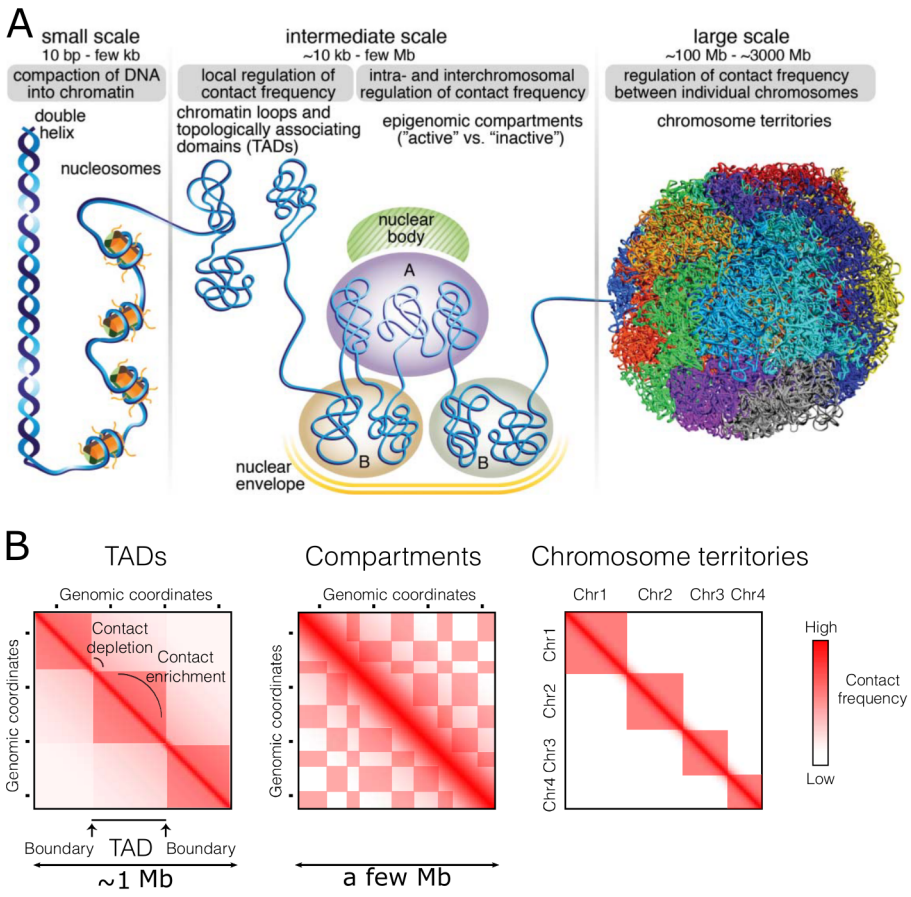


Figure 1: The genome is functionally non-randomly folded at multiple scales in the nucleus. A: The DNA double helix molecule is wrapped around nucleosomes. These histone octamers can be submitted to covalent chemical bond modifications. The epigenetic modifications creating the histone code regulate locally the chromatin environment and gene expression. At scales of a few hundreds of kilobases, chromatin loops and Topologically Associating Domains (TADs) are observed. Loops and TADs exhibit higher frequencies of contacts within them than outside their boundaries. At few Mb, epigenomic compartments organize chromatin in functional compartments that segregate from each other. The A compartment is a transcriptionally active compartment, while the B compartment is repressive and tends to be located at the nuclear lamina. At the scale of the nucleus, chromosomes segregate from each other and preferentially occupy specific radial positions in the nucleus. Reproduced from [19]. **B:** Genome organization as observed in contact maps. TADs form squares of enriched contacts along the main diagonal (left). Compartments exhibit a checkerboard pattern (middle). Chromosomes tend to segregate from each other, low levels of interactions are observed between chromosomes (right). Adapted from [20].

1.2 Hi-C describes genome architecture

1.2.1 Mapping genome-wide 3D spatial contacts using Hi-C

The study of 3D genome organization is based on methods assessing the frequency of physical contacts between different fragments of chromatin. These 3C techniques allow to map fragments of chromatin that are in close proximity in space. A variety of 3C methods were developed to map spatial contacts but most of them are based on the same principle, detailed below. First, chromatin crosslinking chemically creates bonds between two (or more) fragments of chromatin that are in close proximity. DNA is then digested with one (or more) restriction enzymes to fragment bonded DNA segments. These crosslinking and fragmentation steps are key for the downstream analysis of Hi-C data[21]. DNA ends are filled and labelled with biotin for future purification. Finally, ligation between fragments is performed in dilute conditions to favor intramolecular ligation and biotinylated fragments are purified (**Figure 2A**)[22,23]. High-throughput sequencing is then used to map all spatial contacts genome-wide and create the Hi-C contact maps (**Figure 2B**).

Over time, the 3C-related methods expanded from the analysis of one *vs* one fragment of chromatin[24] to genome-wide mapping of contacts with Hi-C[25], and at high genomic resolution with Micro-C[26–28], enabling to inspect chromatin organization at a wide range of scales (from base-pair resolution[29] to chromosomal interactions).

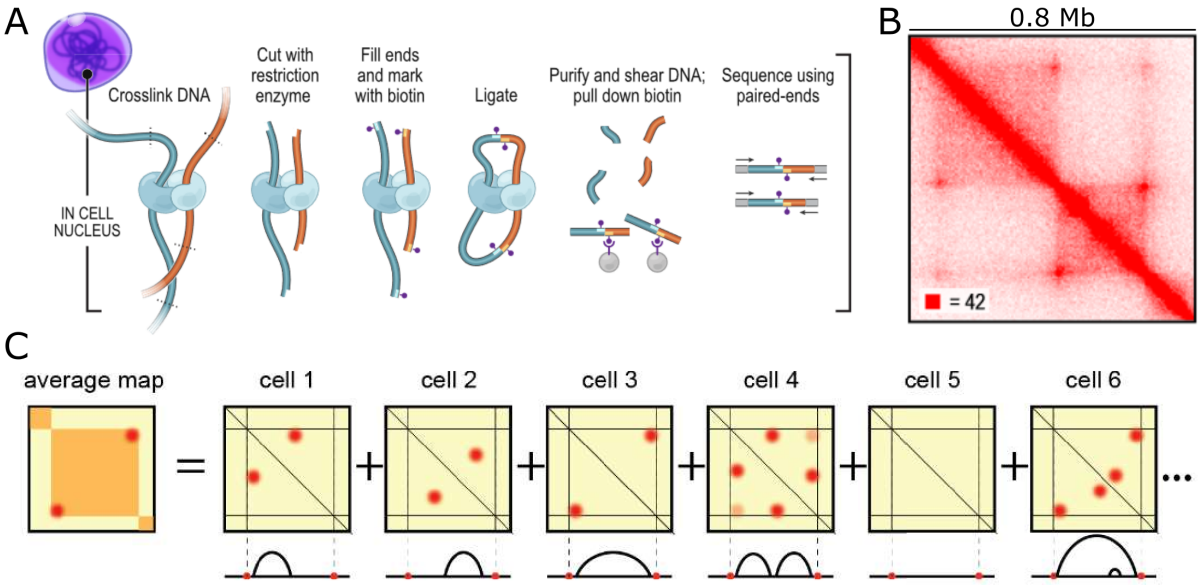


Figure 2: Hi-C maps genome-wide DNA spatial contacts. A: *In situ* Hi-C protocol scheme. Reproduced from [30]. **B:** Example Hi-C map of a 0.8 Mb region in HCT116 cells. Reproduced from [31]. **C:** Schematic showing how a Hi-C map (left) is obtained from the averaging of multiple contacts obtained in many different single cells. The pattern of the average map (left) only emerges from the statistical averaging of preferential contacts within the region in single cells (right). Not all single cells exhibit the contacts observed in the averaged map. Reproduced from [32].

1.2.2 Limitations of Hi-C and other 3C methods

Hi-C uses fixed cells and is generally a bulk average technique where ~1 million of nuclei are used to produce the final Hi-C map. Contacts seen in bulk Hi-C maps (**Figure 2B**) result from the statistical average of ~1 million of cells (**Figure 2C**). Hence, they do not reflect the chromatin architecture of single cells but rather emerge from the averaged preferential interactions between specific sequences (**Figure 2C**)[19]. Moreover, the sparsity of long-range compared to short-range contacts needs to be taken into account during contact map normalization[33]. Also, the consequences of the fixation step in Hi-C are not fully understood: (i) crosslinking occurs between proteins or between proteins and DNA, but not between DNA fragments only, although histone proteins decorate most of the chromatin fiber, to what extent the crosslinking rate is heterogenous genome-wide is unknown; (ii) the distance length at which crosslinking captures contacts is not well defined[34–36] and more importantly, it is unclear if it is constant across the whole genome; (iii) formaldehyde is used as a fixative agent but it is a highly denaturing agent for nucleic acids[37], and using different fixative agents was found to affect downstream analysis of Hi-C maps[21]. As the resolution of

3C methods increases and reaches base pair resolution[29], defining sources of noise and artifacts might become more and more important in the analysis of the resulting contact maps.

However, other 3C-like methods were developed to alleviate some of the drawbacks of bulk Hi-C. Namely, single-cell Hi-C enabled to identify high cell-to-cell heterogeneity in chromatin architecture[38]. Genome Architecture Mapping (GAM) which does not use chemical fixation and keeps single-cell information recapitulated and extended Hi-C findings[39]. Other methods such as Split-Pool Recognition of interactions by Tag Extension (SPRITE) enabled the identification of multiway-specific interactions, both intra and inter-chromosomal[23,40]. Although each of these techniques have their own specificities (*e.g.* levels of resolution depending on genomic distance), they all recapitulated and extended the previous results obtained from bulk Hi-C data[41].

1.3 DNA and chromatin can be modelled as polymers

1.3.1 DNA is a string of monomers linked by bonds

DNA is made of billions of base pairs that are covalently linked to each other. This structure falls into the definition of a polymer: a molecular structure built from similar units (called monomers) bonded together (**Figure 3A-B**). Polymers have universal properties derived from physics theory, which can be applied to chromatin. Most results from statistical physics assume that the polymer is at equilibrium. Although this assumption is violated by the chromatin fiber (multiple energy-consuming processes simultaneously occur on chromatin), polymer models used at a wide range of scales (from DNA around a single nucleosome to the whole genome) recapitulated the spatio-temporal behavior of chromatin in the nucleus[42–44].

1.3.2 Polymer physics can recapitulate chromatin structure and dynamics

Polymers are made of identical monomers (beads) linked by bonds (**Figure 3A**). Consecutive bonds can rotate, relative to each other, which produces a semi-flexibility property of the whole polymer. At short length scales, the polymer is rigid, while at longer scales the polymer is more and more flexible. Therefore, the polymer can adopt an infinite number of conformations (**Figure 3B**). Statistical physics allows to derive scaling laws from the averaging of many conformations, depending on the number of monomers that constitute them[45].

The simplest polymer model is the ideal chain (or freely jointed chain) where monomers ignore each other (*i.e.* monomers can overlap). In this model, the average end-to-end distance increases as $N^{1/2}$, with N the number of monomers, while the contact probability (P(*s*)) decreases as $s^{-3/2}$, with *s* the distance in monomers (or the genomic distance for DNA)[45]. Other models where excluded volume interactions are added (monomers cannot overlap, *i.e.* the chain is self-avoiding) lead to a higher scaling of the end-to-end distance as $N^{3/5}$. In the first experimental measurements of intrachromosomal P(*s*) curves, a s^{-1} scaling was observed (**Figure 3C**). This was interpreted as evidence for a fractal globule, an out-of-equilibrium model without topological entanglements (absence of knots) (**Figure 3C**)[25,46]. However, the fractal globule model did not capture the entire set of experimental data describing genome architecture. The plateauing of mean squared spatial distance as function of genomic distance due to chromatin confinement in the nucleus and differences in the scaling of *s* between organisms and chromatin states[47,48] could not be reproduced using a unique fractal globule model (**Figure 3D**)[49]. To reconcile these observations, an alternative polymer model (the Strings and Binders Switch (SBS) model) was developed. In this model, a self-avoiding polymer chain is folded by the interactions with freely diffusing binders[50]. This model recapitulated imaging and Hi-C experiments altogether and reproduced, as one of other possible conformations, the s^{-1} scaling observed in Hi-C[49].

Apart from structural estimates, polymer physics also predicts the dynamic behavior of the monomers, which are submitted to random motion due to thermal agitation. Unlike single molecules suspended in a liquid medium, the motion of individual monomers is affected by the links between their neighboring monomers. The Mean Squared Displacement (MSD) as function of time interval is a widely used quantity to describe stochastic motion of a particle. It is defined as: $MSD(\Delta t) = \langle \left(\mathbf{r}(t+\Delta t) - \mathbf{r}(t)\right)^{\mathbf{2}} \rangle$ where $\mathbf{r}(t)$ is the position vector of the particle at time *t*, and brackets denote averaging over time. An isolated diffusing molecule (*i.e.* not part of a chain) will display a MSD proportional to the time interval: $MSD(\Delta t) \propto \Delta t$[42]. By contrast, for an ideal chain, a single monomer of the polymer will undergo sub-diffusion with the MSD increasing like the square root of the time interval: $MSD(\Delta t) \propto (\Delta t)^{0.5}$ over short timescales[42]. It then reaches a plateau at larger time intervals: $MSD(\Delta t) = constant$ if the polymer is confined, well recapitulating *in vivo* tracking of chromatin loci[51–53]. This scaling of $MSD(\Delta t)$ with a powerlaw of exponent 0.5 is characteristic of the Rouse dynamics, a good approximation of chromatin behavior[54].

1.3.3 Molecular dynamics simulations help to understand genome architecture

The statistical properties derived from polymer physics are used to model genome architecture. Using molecular dynamic simulations, the chain formed by linked monomers can be subjected to stochastic motion, such as Langevin dynamics, which approximates the effect of the solvent instead of fully computing the interactions between the polymer and the solvent. The overall conformation of the polymer(s) and its motion can be recorded and further analyzed (**Figure 3B**). This computational approach potentially enables a large variety of assumptions and hypotheses to be tested, at low temporal and economic costs as compared to experiments. From these *in silico* models, various approaches allow to better understand the mechanisms driving genome organization, as well as highlight the remaining unknown or poorly described mechanisms[55,56]. Such approaches can be grouped in two different categories described below: bottom-up and top-down models (**Figure 3E**)[56,57].

Bottom-up approaches: chromatin conformation emerges from polymer physics laws

Bottom-up approaches model chromatin as a polymer obeying the classical physics laws. *A priori* mechanism(s) of polymer folding (*e.g.* interactions, static or dynamic links between beads) are applied on the polymers. The resulting polymer conformations or motions are compared to experimental data or theoretical considerations to find sufficient and simple assumptions recapitulating the observations (**Figure 3E**). These predictive models do not rely on experimental data. Although, parameter space exploration can be needed to identify the most probable parameters, as well as the landscape of possible alternatives, physics laws constrain this search by multiple orders of magnitude as compared to top-down approaches[58]. The output conformations mostly depend on the set of assumed rules, hence rendering their interpretation easier due to their parsimonious nature[32,59–62].

However, the agreement between simulations and experiments is not a proof that the assumed mechanism occurs in the cells. Many different models (even beyond the ones tested) might match experimental data and a single model agreement is not sufficient to claim experimental relevance but can guide future experimental design towards a better understanding of the studied mechanism.

Top-down approaches: chromatin conformation is defined by experimental data

Top-down approaches are data-driven. Based on chromosome conformation capture (3C) experimental methods, a specific genomic region is used to define the interactions and links between polymer beads (**Figure 3E**). Because of the stochastic nature of polymer structures, these models better satisfy the input data (typically Hi-C matrices) when yielding an ensemble of single conformations[63,64] rather than a single consensus conformation[65]. Taking into account physical constraints of the polymer, the best set of modelled parameters is found by finding the minimum of the defined cost function (generally a measure of the disagreement between simulated and experimental contact matrices)[56].

These models proved to be powerful in identifying and explaining the different forces driving the spatial organization of a single locus or whole chromosomes[63,64,66,67]. Moreover, once the interactions between beads are defined, *in silico* perturbations (such as removal of specific genomic sequences or of one specific folding mechanism) of the polymer can be studied to predict the experimental outcome of the corresponding genomic mutation[63,64]. For example, it is possible to silence the interactions from specific bead(s) (corresponding to a specific genomic locus) on the polymer. This *in silico* perturbation reproduces an experimental genomic deletion of the locus. This strategy enabled the identification of a genomic locus controlling long-range interactions within the Tsix locus on the X chromosome[63]. However, due to the data-based approach, top-down models heavily depend on the initial quality of experimental data and are sensitive to bias or overfitting for machine learning methods[68]. Therefore, they may identify less generalizable mechanisms than bottom-up approaches since they usually focus on a specific genomic region[63,64].

The chromatin fiber can be modelled as a polymer at various scales (from a few nucleosomes to the whole genome). Based on polymer physics theory, polymer simulations and comparison with experimental data, hypothesized molecular mechanisms of genome folding can be tested.

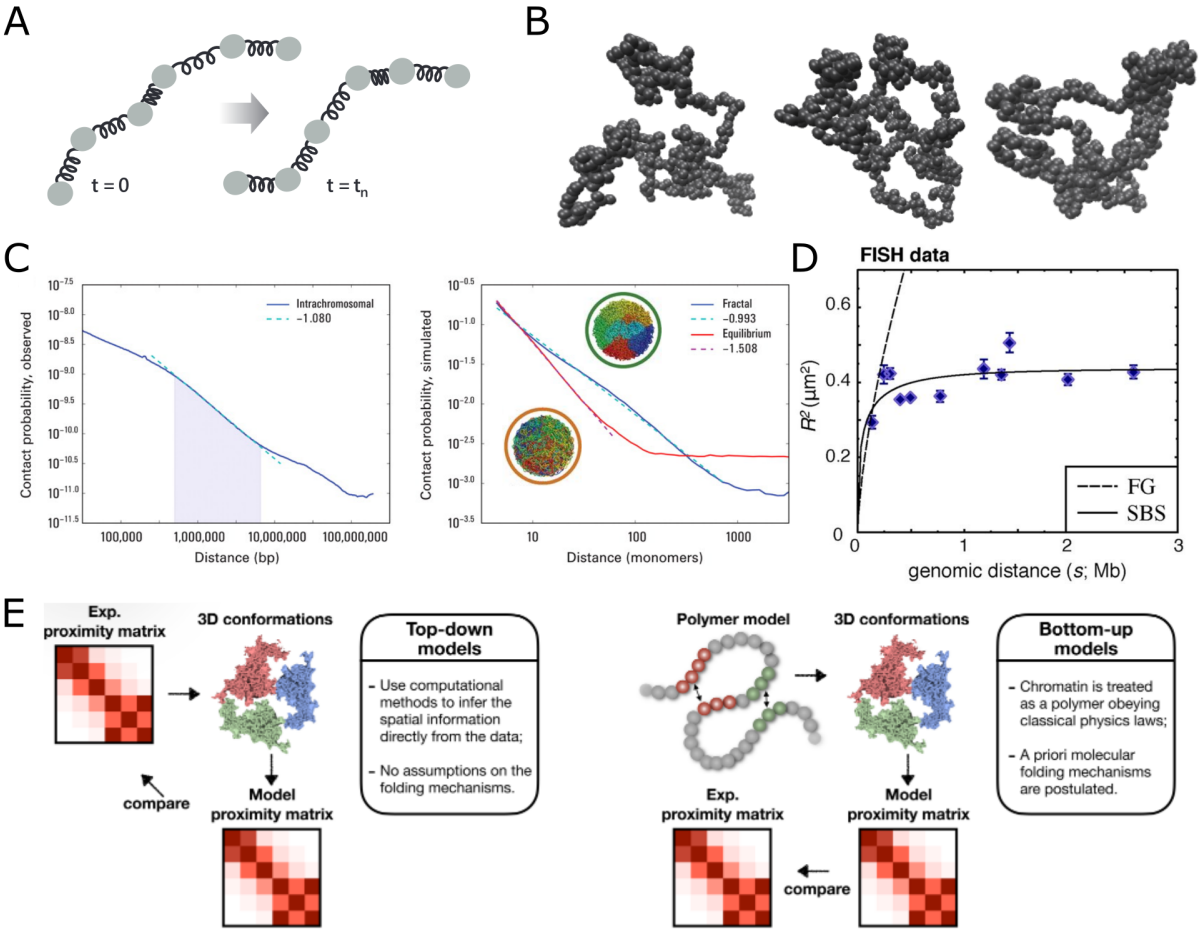


Figure 3: Polymer simulations are used to model chromatin. A: Polymers are made of beads linked by bonds acting as springs. The motion of monomers, as well as topological properties of the polymer chain can be derived from statistical polymer physics. Reproduced from [69]. **B:** Example of three different polymer conformations. During the course of a simulation, the polymer is subjected to stochastic Langevin dynamics. These random movements can produce an infinite number of polymer conformations. The polymer conformations can be recorded as function of time and compared to theoretical considerations or experimental data. **C:** Contact probability as function of genomic distance measured experimentally by Hi-C (left, averaged over the entire genome) and predicted from an equilibrium or fractal globule model (right). Examples of simulated chromosome conformations are shown in the right panel (green circle is fractal globule and orange circle is the equilibrium model). Adapted from [25,70]. **D:** Mean squared spatial distances as function of genomic distance obtained by DNA FISH. A plateau is seen starting at ~1 Mb. Predictions from the fractal globule (FG) and Strings and Binders Switch (SBS) models are shown. Reproduced from [49]. **E:** Two different approaches can be used to compare polymer simulations and experimental results: top-down models (left) where experimental contact data are used to infer polymer conformations, and bottom-up models (right), which are based on polymer physics laws and hypothesized molecular mechanisms. Reproduced from [56].

1.4 The genome is spatially non-randomly organized

1.4.1 Chromosomes occupy specific territories in the nucleus

Despite the large size of chromosomes, they do not exhibit a homogeneous spatial distribution in the nucleus. Indeed, chromosomes tend to occupy preferential radial positions within the nucleus, named chromosome territories (**Figure 1**)[71–73].

The appearance of chromosome territories might be explained by the timescale needed for chromosomes to relax after mitosis. Due to the inability of polymer fibers to cross each other, the relaxation time needed to reach genome-wide equilibrium is much longer than the human cell cycle duration[60]. This prediction is supported by the agreement of the experimental contact probability scaling s^{-1} with an out-of-equilibrium fractal globule model, where chromosomes occupy distinct locations in the nucleus (**Figure 3C**)[25]. However, active and passive mechanisms (*e.g.* topoisomerases, differential transcriptional activity, compartmentalization, loop extrusion) occurring on the genome might help to decrease this relaxation time following mitosis and were not taken into account in these simulations[69].

1.4.2 Chromosome compartmentalization into active and inactive domains: a functional partitioning of the genome

Electron microscopy images of interphase nuclei already exhibited spatial partitioning of the genome. Dense chromatin regions near the nuclear envelope and around nucleoli were identified as heterochromatin. By contrast, the center of the nucleus was occupied by less dense chromatin, the euchromatin[74].

Hi-C maps show the frequency of contacts between each portion of the genome *vs* all other portions of the genomes. Obviously, a strong main diagonal appears as DNA segments close on the DNA sequence are also close in space (**Figure 1B** and **Figure 2B**). But surprisingly, specific patterns emerged from the main diagonal, which were not expected if the chromatin would be randomly organized in the nucleus.

Specifically, chromatin compartments on the order of a few Mb that segregate from each other were identified as a checkerboard pattern in Hi-C maps (**Figure 4A** and **Figure 1B**)[25]. An eigenvector analysis, together with ChIP-Seq analysis, suggested that the two compartments corresponded to transcriptionally active chromatin (A compartment or

euchromatin, defined by permissive histone marks and open chromatin) and a repressive compartment (B compartment or heterochromatin, defined by repressive histone marks and dense chromatin regions)[75]. A deeper analysis of Hi-C data, together with polymer simulations, suggested that A and B compartments could be separated into sub-compartments[30,76,77]. These sub-compartments have varying degrees of transcriptional activity, as well as histone epigenetic histone marks and could be stemming from different mechanisms of segregation (lamin-associated heterochromatin, preferential *cis* interactions, replication timing)[76].

Compartment location correlates with functional processes of the genome, such as a later replication timing in B compartments, and a higher gene expression level in A compartments[69]. However, it is difficult to disentangle the causes and consequences from these correlations. Nevertheless, compartments were found to be remodeled upon cell differentiation[78] and are sufficient to cluster cell lineages[79], suggesting a functional role associated with genome compartmentalization.

The mechanism of compartmentalization is not yet well understood. However, polymer simulations, based on the observation that A and B compartments are enriched for specific histone marks, helped to define a compartmentalization mechanism. The sole addition of weak specific attractive interactions between monomers of the same type in heteropolymers (polymers where monomers exhibit different properties) is sufficient to mediate compartmentalization as seen in Hi-C maps (**Figure 4B-C**)[61,77,80,81]. More precisely, compartmentalization is mostly driven by stronger attractions between the B regions as compared to weaker attractions between A regions[81]. The molecular mechanism by which this attraction happens could emerge from phase separation[44,82,83], for instance in the B compartment, driven by HP1α's ability to self-interact[84,85].

Thus, A and B compartments are more likely to interact with regions of the same compartment than with regions of the other compartment. These specific attractive interactions are sufficient to explain the segregation between the two compartments.

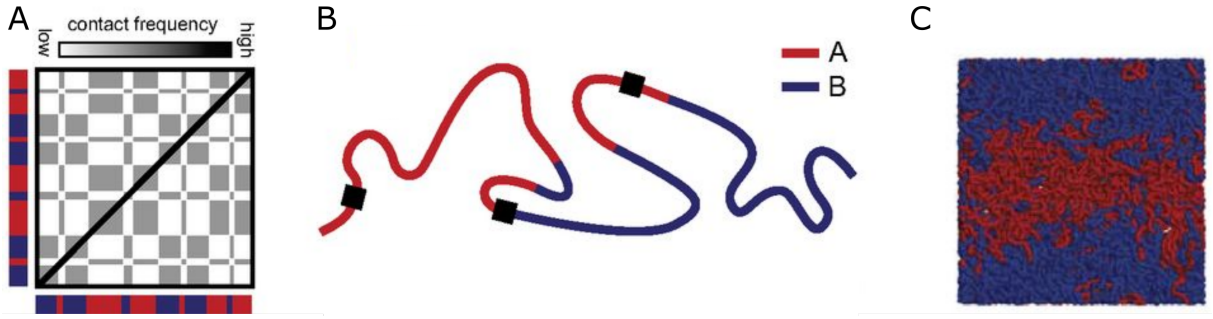


Figure 4: Compartmentalization can be modelled by the sole addition of interactions between regions of the same compartment. A: Schematic showing compartments as observed in Hi-C maps (A compartment in red and B compartment in blue, around the Hi-C map). **B:** Heteropolymer with A and B regions. Interactions within the A compartment (A) and within the B compartment (blue) are sufficient to mediate compartmentalization. **C:** Example result of a polymer simulation showing spatial segregation of A and B compartments. Panels A-C are reproduced from [61].

1.4.3 Topologically associating domains and chromatin loops preferentially interact within themselves

As Hi-C genomic resolution increased, more specific structures could be discerned from the contact maps. At the scale of a few hundred kilobases (kb) (an intermediate distance scale whose structure remained poorly characterized), squares of increased contact frequencies, named Topologically Associating Domains (TADs), were uncovered. TADs display more interactions within their domain than with regions located outside their boundaries (**Figure 1B**)[86,87]. The corner of TADs (which defines the TAD boundaries, also called borders or anchors) can exhibit a peak in interaction frequencies (**Figure 2B**). These peaks indicate the presence of loop structures where the two distant anchors exhibit a higher contact frequency than with the interior of the loop. Visual inspection of Hi-C maps at the resolution of TADs shows that these structures cover almost the entire genome. Automatic calling of TADs and loops show that there are several dozens of thousands of these structures with a median size of about 200 kb throughout the genome[88]. These structures were later observed by imaging as forming physical domains in single cells[89,90].

In addition to the contact enrichment peaks defining loops, other specific features were uncovered at TAD location in Hi-C maps. Stripes (or flames), which are horizontal or vertical lines darkening the TAD edges, indicate preferential contacts between the TAD boundary and the whole TAD compared to other loci within the TAD (**Figure 6A**)[91]. Moreover, a complex

network of TADs and loops emerged from high genomic resolution Hi-C when it was noticed that smaller TADs could be nested within bigger TADs (**Figure 6A**)[26,27].

Hi-C revealed the presence of chromatin loops and TADs spanning a few hundreds of kb. These structures interact more within themselves than with other parts of the genome and create a complex network of nested chromatin organization. Once these structures were unveiled, two main questions emerged:

- What is the molecular mechanism of TAD formation?
- What is the biological role(s) of these structures?

We will discuss these two questions in the following sections.

2. The loop extrusion model: cohesin extrudes loops between CTCF anchors

2.1 CTCF-bound CTCF motifs and cohesin are found at loop anchors

As the location of chromatin loops was identified genome-wide, searching for potential protein candidates involved in their formation became possible. The inspection of genome-wide ChIP-Seq data revealed that cohesin subunits and the CCCTC-Binding Factor (CTCF) were associated at 86% of loop anchors[30]. Moreover, examining the DNA sequences located at loop anchors unveiled the presence of CTCF binding sites oriented more than 90% of the cases in a convergent orientation[30]. Finally, cohesin and CTCF were found to colocalize spatially as assessed by super-resolution imaging[92,93] and ChIP-Seq[30], and they co-immunoprecipitated together[92,94]. These observations made CTCF and cohesin ideal candidates for putative proteins involved in loop formation.

2.2 Depletion of cohesin or CTCF differently alters Hi-C maps

To investigate further the role of these proteins in chromatin loop formation, Hi-C maps of cells depleted from CTCF or cohesin subunits were built.

The depletion of cohesin led to a drastic change in Hi-C maps: the elimination of all loops throughout the genome[31]. TADs and loops observed in wild type (WT) cells vanished within 1 hour after cohesin depletion. In these cohesin-depleted Hi-C maps, no specific feature other than the typical genomic distance-dependent decay in contact frequency was distinguishable at the scale of TADs and loops (**Figure 5A**). After allowing the cells to synthesize cohesin again, chromatin loops rapidly reappeared within 90 min (**Figure 5B**), proving the formation of such loops was cohesin-dependent[31].

The depletion of CTCF led to more subtle changes in genome-wide Hi-C maps. TAD and loop boundaries were still detectable but were faded (**Figure 5C**)[95,96]. This suggested that CTCF is not directly involved in the formation of loops, but rather defines their boundaries.

These experimental observations suggested that cohesin is the main protein involved in the formation of chromatin loops, while CTCF defines their boundaries. Further investigation led to the loop extrusion model as the mechanism explaining chromatin loop formation.

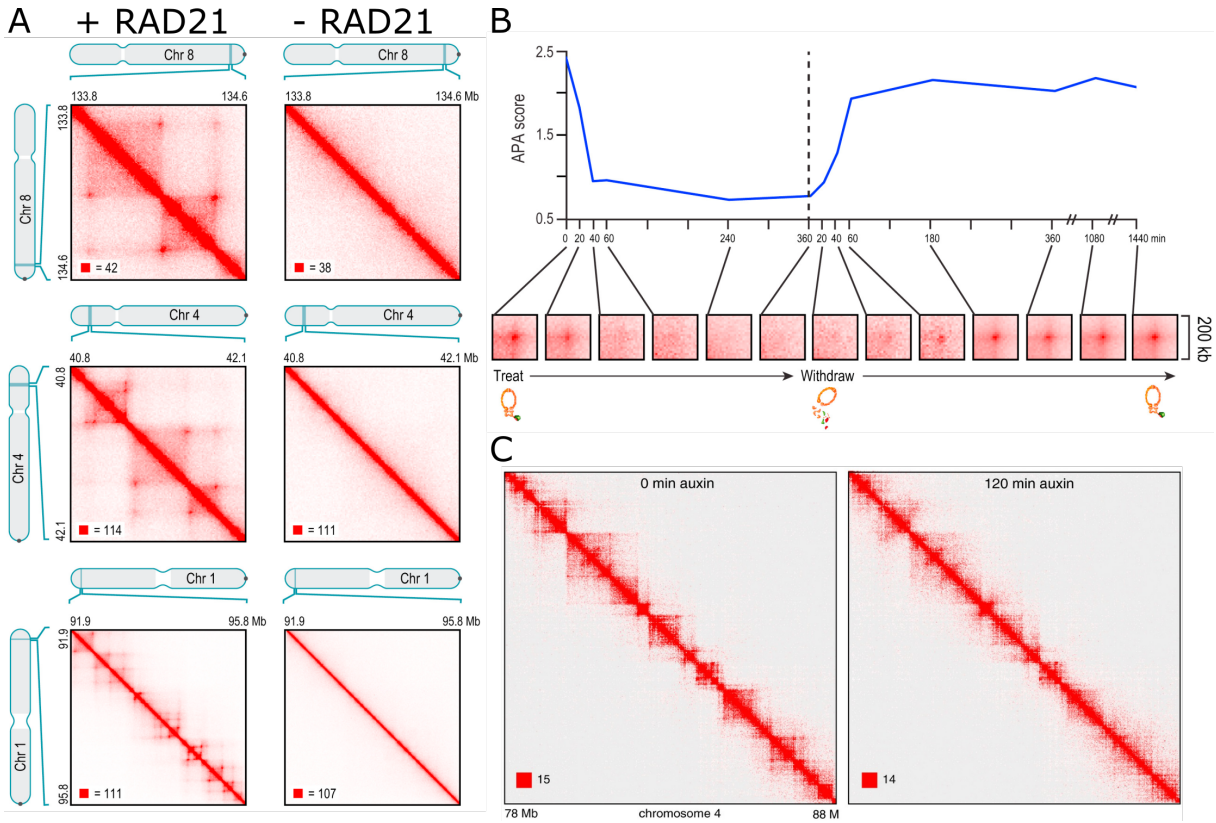


Figure 5 : Depletion of cohesin and CTCF led to different changes in Hi-C maps. A: The depletion of RAD21 (subunit of the cohesin complex) led to the complete disappearance of loops and TADs. **B:** Pileups of Hi-C maps centered on chromatin loop anchors genome-wide after RAD21 depletion (treat), and after allowing for cohesin resynthesis (withdraw). Loops disappear within one hour of treatment and reappear after 90 min of cohesin resynthesis. The Aggregate Peak Analysis (APA) score is a measure of the intensity of dots in pileups Hi-C matrices. **A** and **B** are reproduced from [31]. **C:** Hi-C maps without (left) and with (right) CTCF depletion. After CTCF depletion, TADs are still detectable, but their boundaries are faded. Reproduced from [95].

2.3 Cohesin is the motor of loop extrusion and is halted by CTCF

Because random collisions due solely to thermal fluctuations of chromatin cannot explain the contact peaks observed in statistically averaged Hi-C maps, a molecular mechanism was needed to explain the formation of loops[97]. The loop extrusion model recapitulated the experimental observations made from Hi-C. According to this model, cohesin binds on chromatin and forms a loop by actively pulling out (extruding) DNA, until it unbinds or encounters an obstacle such as a bound convergent CTCF-site (**Figure 6B**)[98]. If cohesin reaches a bound CTCF site, cohesin stalls. If CTCF unbinds, cohesin can resume extrusion or unbinds, which dissolves the loop[99]. CTCF binding site orientation defines the

boundaries of TADs where cohesins stop extruding, creating a peak of contact probabilities in Hi-C maps.

Polymer simulations successfully recapitulated experimental observations by modelling the loop extrusion process. Polymer models where an extruding complex (or extruder) dynamically extrudes loops until encountering an obstacle defining the TAD anchors were sufficient to accurately recapitulate experimental Hi-C data[32,62]. These models matched the genomic distance-dependent decay in contact frequencies observed in Hi-C data. They also displayed the features observed in Hi-C such as TADs, loops and stripes (**Figure 6A**). Not only did these models recapitulate the experimental observations, but they also predicted the effect of mutated boundaries, as well as enabled estimations of dynamic parameters of loop extrusion. As expected by the loop extrusion model, deletions or inversion of the CTCF sites at TAD boundaries led to fusion of neighboring TADs and disappearance of the peak at the deleted boundary, both in polymer simulations and in experiments[62,100]. Finally, the extruder processivity (average size of a loop extruded by a single cohesin during a single binding event) and separation (genomic distance between two bound extruders) could be estimated by exploration of the parameter space. The extruder processivity and separation in simulations led to estimates of about 120 kb for both quantities[32].

The loop extrusion model well explains features observed in Hi-C. Focal peak enrichments are created by cohesin stopping at specific locations such as CTCF anchors. Stripes (or 'flames') can be explained by unidirectional extrusion from a TAD boundary[91,101], while 'jets' or 'flares' (antidiagonal structures observed in quiescent cells) might emerge from bidirectional extrusion at specific cohesin loading sites[102,103] (**Figure 6A**). This model was further verified by numerous perturbation studies (described in section **3. Loop extrusion dynamics is tightly regulated**) and a more detailed view of the molecular process was described[98,104].

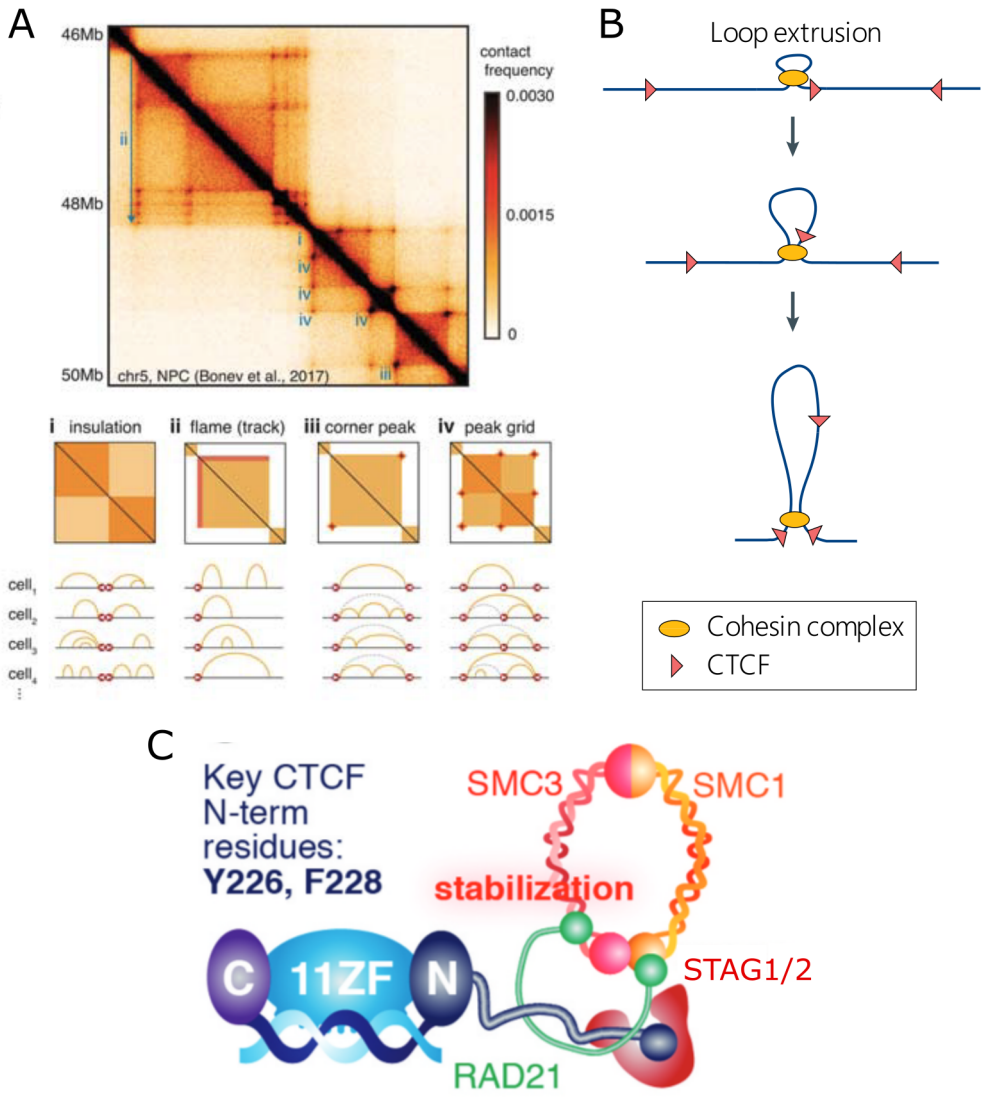


Figure 6: The loop extrusion model: cohesin extrudes loops and halts at bound-CTCF sites. A: Hi-C features (top and schematized in the middle) are well explained by a dynamic loop extrusion model. Several single-cell models of interacting regions are shown at the bottom for each Hi-C feature. Reproduced from [105]. **B:** In the loop extrusion model, the cohesin complex extrudes a loop until encountering a pair of convergent CTCF-bound CTCF sites. Reproduced from [98]. **C:** The cohesin ring (right) is composed of several proteins: RAD21, STAG1 or STAG2, SMC1 and SMC3. SMC1 and SMC3 form the cohesin ring with their coiled-coil regions and join at the hinge (top). CTCF halts cohesin when its N-terminal domain faces cohesin, and this protein-protein contact might stabilize CTCF and cohesin interaction. Adapted from [99].

2.4 Cohesin: a ring-shaped protein complex extruding loops

Human cohesin is a ring-shaped protein complex consisting of four subunits: SMC1, SMC3, RAD21 and either STAG1 or STAG2. The SMC subunits are coiled coil domains of about 50 nm long, which form the two sides of the cohesin ring (**Figure 6C**)[98].

2.4.1 Cohesin DNA-loading mechanism upon binding

The exact molecular mechanism of cohesin loading is not yet clear. Two possible scenarios are: (i) topological loading (DNA is inserted inside the cohesin ring by opening of the hinge, **Figure 6C** and **Figure 7A**) or (ii) pseudo or non-topological loading (the loop is extruded without the need for cohesin to open its ring) (**Figure 7A**)[104]. It was shown that cohesin is able to extrude loops by pseudo or non-topological loading *in vitro*[106]. Moreover, the observation that Structural Maintenance of Chromosomes (SMC) complexes are able to traverse physical roadblocks (up to 200 nm) much bigger than their ring size (50 nm) argues in favor of non-topological loop extrusion loading rather than topological or pseudo-topological[107,108]. These two scenarios are not mutually exclusive and could co-exist in the cells[109]. They may be used differentially to achieve the different cohesin functions (*e.g.* pseudo or non-topological loading for loop extrusion in G1 and topological loading for sister chromatid cohesion[98]).

Although several evidence argue in favor of the non-topological mechanism of cohesin chromatin binding, there is still no clear evidence to reject one loading mechanism.

2.4.2 Cohesin effectively extrudes loops bidirectionally

After cohesin loading on DNA, loop extrusion starts. Loop extrusion could happen asymmetrically (a single DNA strand is reeled into the cohesin ring) or symmetrically (both DNA strands are reeled into the cohesin ring) (**Figure 7B**). Polymer simulations indicate that asymmetric extrusion alone does not meet the required compaction observed in mitotic genome organization[110–112]. However, models of effective bidirectional extrusion such as switching asymmetric extrusion or a mix of one-sided and two-sided extrusion, as well as purely two-sided extrusion are able to recapitulate experimental and theoretical findings (**Figure 7B**)[110–112]. *In vitro* visualization of loop extrusion with purified components showed that human cohesin extrudes loops symmetrically[106,113]. Evidence for bidirectional extrusion was also present in Hi-C maps of quiescent cells where 'jets' (antidiagonal stripes) were observed[102,114]. Thus, cohesin can effectively extrude bidirectionally. However, this may be achieved by cohesin dimers, each extruding unidirectionally or by cohesin monomers extruding bidirectionally (**Figure 7C**).

The stoichiometry of cohesin complexes needed for loop extrusion remains under debate[115]. Cohesin was found to exist mostly as a monomer in living yeast[116] and human cohesin monomers are sufficient to extrude loops *in vitro*[106,113]. However, dimerization or oligomerization of cohesin is also observed *in vitro*[113] and *in vivo* (>8% of cohesin molecules exist in an oligomerized form)[117]. Thus, the minimal functional unit of loop extrusion is a monomeric cohesin but whether cohesin is extruding loops as a monomer or as dimers (or oligomers) *in vivo* remains unclear.

2.4.3 Cohesin extrudes loops at high speed

In vitro studies of loop extrusion on naked and nucleosomal DNA showed that cohesin-dependent loop extrusion was an active process, consuming ATP (~1 ATP molecule per second)[106,113]. The studies could estimate the loop extrusion speed at 0.5-1 kb/s[106,113]. This speed is moreover tension-dependent and is lowered in high-tension conditions[113,118]. Although this estimate allows to better understand the kinetics of loop formation, it has not been measured *in vivo* where other components might accelerate or decelerate cohesin-mediated loop extrusion[104,107]. For instance, DNA tension can be created by several nuclear processes (*e.g.* transcription or replication forks). Although suffering from approximations, indirect measures of loop extrusion speed *in vivo* estimated the process at a lower rate of 0.1-0.2 kb/s (from the average size of TADs: 200 kb and the mean binding duration of cohesin: 20 min[92])[119] or about ~0.38 kb/s from the reappearance of TADs after cohesin depletion and resynthesis[31]. In addition, the *in vivo* extrusion speed might not be homogenous throughout the genome. Indeed, loops in the A compartment were found to reappear faster than loops in the B compartment after cohesin depletion and resynthesis[31]. This difference could stem from differences in loop extrusion speed, but also from differential binding of cohesin (or other factors) in A and B compartments.

Cohesin is a fast nuclear motor since these estimates of loop extrusion speed are at least an order of magnitude higher than the speed of some nuclear motors (*e.g.* 0.02 kb/s for RNA Pol II[120] and 0.025 kb/s for the RSC translocase[121]).

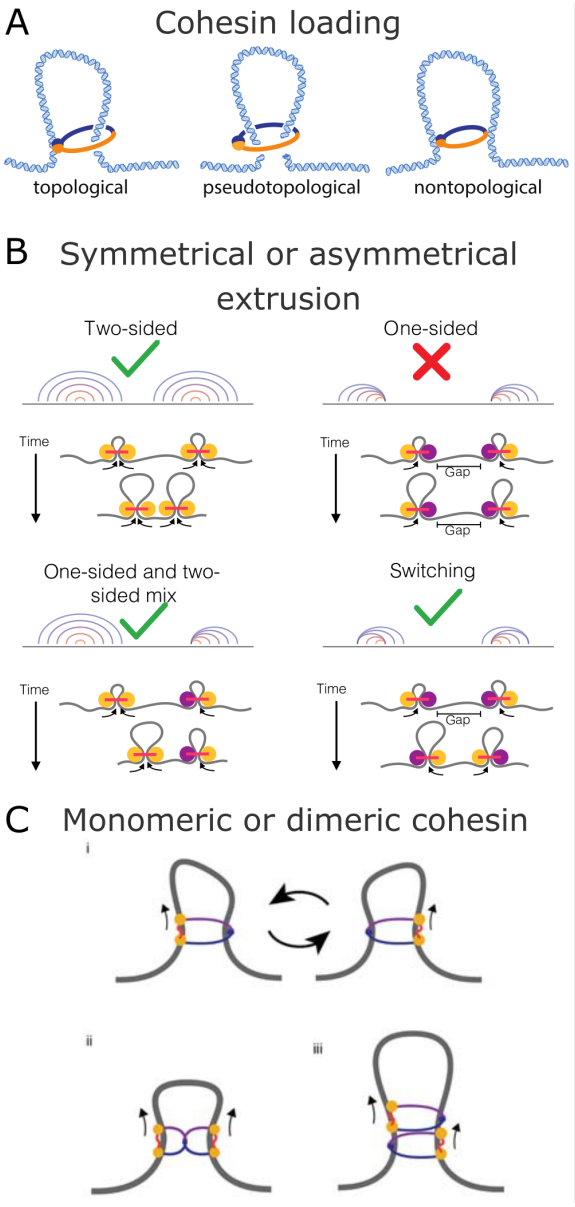


Figure 7: Models of loop extrusion loading, extrusion directionality and cohesin stoichiometry. A: Several possible mechanisms of cohesin loading on DNA. In the topological model, DNA is entrapped within the cohesin ring, unlike in the two other mechanisms. Pseudo or nontopological mechanisms are the most likely models for cohesin-mediated loop extrusion. Reproduced from [107]. **B:** Cohesin could extrude unidirectionally (also called one-sided or asymmetrically) or bidirectionally (also called two-sided or symmetrically). For each mechanism, its ability to recapitulate or not theoretical both interphase and mitotic considerations is indicated by a green check mark or a red cross symbol, respectively. Adapted from [111]. **C:** Possible models of cohesin extrusion with monomeric (i) or dimeric ((ii) and (iii)) cohesin complexes. Cohesin can perform loop extrusion in its monomeric or dimeric (and oligomeric) states *in vitro*. Reproduced from [105].

2.4.4 Cohesin trafficking on a crowded chromatin fiber

The DNA molecule in nuclei is bound by multiple proteins. Nucleosomes wrap DNA around them, the transcription machinery as well as transcription factors are continuously binding and unbinding from DNA. Moreover, condensates as well as spatially distinct chromatin bodies (*e.g.* nucleoli) locally increase the concentration of proteins[122]. Despite the *in vitro* characterization of loop extrusion, it is not *a priori* obvious to what extent chromatin crowding affects loop extrusion. Several pieces of evidence suggest that cohesin can cope with chromatin crowding. First, cohesin molecules can traverse one another *in vitro* and thus create Z-loops (a double loop created by two cohesin molecules passing through each other) while simultaneously extruding the same DNA molecule[123]. *In vivo* assessment of bacterial Hi-C patterns showed that SMC complexes can indeed traverse each other[124]. This showed that cohesin extrusion on chromatin should not create traffic jams. Second, the interaction of cohesin with other molecules was assessed *in vitro*. Sparsely DNA-bound nucleosomes[113] and other single DNA-bound proteins were not found to block cohesins more than 10% of the time[107] but can restrict its movement[125]. Furthermore, synthetic beads bigger (up to 200 nm) than the cohesin ring size (50 nm) allow cohesin to extrude loops (about 50% of cohesins pass through the obstacle). Although these particles can transiently trigger cohesin pausing, extrusion can resume after stalling.

Thus, cohesin is a ring-shaped protein complex that can bidirectionally and/or unidirectionally extrude loops at high speeds and bypass obstacles bigger than its own size. Because cohesin can bypass large obstacles, a specific barrier is required to halt extrusion and defines TAD borders. Despite its small size (3-5 nm), CTCF specifically halts extruding cohesin molecules[99].

2.5 CTCF is an oriented permeable boundary to extruding cohesins

2.5.1 CTCF specific orientation halts cohesin

From the identification of CTCF sites at chromatin loop anchors, a striking feature was the convergent orientation of CTCF sites. CTCF sites at loop boundaries were generally found (90%) in a convergent orientation[31]. A careful investigation of CTCF structure as well as functional assays shed light on the polarity mechanism by which the 5 nm CTCF protein halts extruding cohesins (**Figure 6B**).

Since the convergent orientation of CTCF sites seemed to be important for CTCF-mediated cohesin halting, inversions of CTCF sites were specifically made at TAD borders. As expected, these genetic modifications of CTCF sites led to a decrease in contact frequency and in cohesin binding at the inversed CTCF sites[62,126–128]. Hence, the molecular polarity of CTCF is key to its ability to halt extruding cohesins.

CTCF is a protein containing 11 zinc finger domains (**Figure 6C**). These domains were identified to bind DNA (ZF 3-9) or RNA (ZF1, 10-11)[129]. Moreover, the *C*- and *N*-terminal domains contain sites of post-translational modifications such as parylation or sumoylation[129,130]. Systematically deleting the *C*- or *N*-terminal domains from CTCF transgenes allowed to define how the polarity of CTCF binding halts cohesin. The deletion of CTCF *N*-terminal domain blurred the frontiers of TAD and loop domains, similarly to a CTCF depletion, and is therefore involved in cohesin positioning at CTCF sites[130]. Further structural work confirmed that CTCF interacts with cohesin by its *N*-terminal domain and suggested that this interaction stabilizes cohesin by preventing its unbinding from DNA[94]. Two point mutations in the CTCF *N*-terminal domain were sufficient to blur TAD boundaries in Hi-C (recapitulating whole CTCF depletion Hi-C maps, **Figure 5C**[95,96]) and to decrease cohesin binding at CTCF sites[94].

Although depletion of CTCF does not affect the overall amount of chromatin-bound cohesin[130,131], the molecular mechanism by which CTCF halts cohesin remains unclear[99]. Identification of the two key CTCF residues interacting with cohesin proved that a direct protein-protein interaction occurs between cohesin and CTCF, albeit with low affinity ($K_d = 0.6$ µM)[94]. Since these residues bind the same cohesin interface as the cohesin unloader Wings Apart-Like Protein (WAPL), CTCF might help stabilizing cohesin once it reached a CTCF site by competing with WAPL (**Figure 9C**)[99].

Thus, convergent bound CTCF sites position CTCF to face its *N*-terminal domain towards incoming extruding cohesins and explains the molecular polarity of CTCF barriers. Nonetheless, CTCF is not an impassable obstacle and extruding cohesins can sometimes bypass CTCF-bound CTCF motifs.

2.5.2 CTCF is a permeable obstacle to loop extrusion

Although CTCF has the ability to block cohesin, multiple evidence show that it is not a completely impermeable barrier. Single-cell imaging of chromosome organization revealed a high variability in chromatin organization and TAD structures at the single cell level[89,132–135]. While averaged spatial distance-maps recapitulated genome-wide Hi-C maps, single-cell distance maps indicated that TAD structures exist at the single-cell level but that their boundaries are less defined than in statistically averaged maps[59,133]. This observation highlighted that CTCF boundaries were not as impermeable to cohesin as averaged Hi-C maps suggested. High-resolution Hi-C (Micro-C) also indicated that CTCF sites can be bypassed. Indeed, arrays of peaks aligned on stripes in Micro-C maps suggested that cohesin can bypass a CTCF site and reach further CTCF sites[26,27]. A deeper analysis of Hi-C data also indicated that the borders between TADs were often not sharp but rather displayed a zone of transition of elevated contact frequency[136], which is expected if cohesin can bypass a CTCF border. Finally, the permeability of CTCF boundary was estimated *in vitro*: CTCF facing cohesins on their *N*-terminal side had a 75% chance of blocking cohesin at their first encounter[137].

The insulation created by cohesin-mediated loop extrusion inside TADs is only about a 2-fold enrichment in contact frequency compared to outside TADs[136,138]. Indeed, consecutive TADs can significantly overlap in space despite the presence of CTCF sites borders[89,139]. This domain intermingling is also observed as increased contact frequency zones at TAD borders in Hi-C maps[136]. Loop extrusion may enable consecutive TADs to intermingle to a relative extent by stacking TADs borders together in space[140]. CTCF borders are leaky and can allow cohesin to travel between consecutive TADs and increase their spatial overlap.

Thus, although TAD contact frequencies are enriched within themselves relative to outside their boundaries, CTCF is not an impermeable barrier to cohesin. This property could promote intermingling between consecutive domain rather than pure insulation. However, CTCF is not the only protein that can halt cohesin-mediated extrusion and other factors were identified as loop extrusion barriers.

2.6 CTCF is not the only barrier to cohesin

Despite the high observed colocalization of cohesin and CTCF in the nucleus[92,93] and at TAD borders[30], CTCF is not the only protein that can stop cohesin extrusion. Approximately 20% of borders do not depend on CTCF, indicating that other factors can act as loop extrusion barriers[141].

2.6.1 RNA polymerases are mobile barriers to loop extrusion

In CTCF-depleted cells (absence of the main cohesin barrier), cohesin was found to accumulate at transcription start sites of active genes[142,143]. In addition, transcription can relocate cohesin over long distances (up to 70 kb) and actively transcribing genes are freed from cohesin[142,144]. Both experimental and polymer simulations showed that RNA Pol II is able to act as a moving barrier for SMC complexes and translocate extruding complexes in bacterial cells. RNA Pol II can also reduce SMC complex speed if both protein complexes move in opposite directions[145]. This relocation of extruding cohesin by transcribing RNA Pol II was shown to be transcription-dependent in mammalian cells[142,143]. Finally, inhibition of transcription affects cohesin distribution in the whole nucleus, maybe due to a lower supercoiling in transcription-depleted cells[146].

Complicating the picture, an RNA binding domain of 38 amino acids was identified in CTCF. The loss of this domain led to loss of chromatin loops normally maintained by cohesin/CTCF, suggesting a supplementary role of transcription (or the RNA molecule itself) in the structural shaping of chromosome organization[147].

2.6.2 Other proteins can interact with cohesin and affect loop extrusion

Unveiling that CTCF was not the sole barrier to loop extrusion and refining the loop extrusion model using mobile barriers revealed that other proteins could finetune loop extrusion translocation.

For instance, the minichromosome maintenance (MCM) complex, essential for DNA replication, was found to be a permeable barrier to loop extrusion based on *in silico*, *in vitro* and *in vivo* data[148]. It is not surprising to identify a DNA replication machinery protein as a barrier since stalled replication forks were identified as barriers for loop extruding

cohesins[149]. Also, the border of TADs were found to correlate with replication domain boundaries[150], strengthening the putative interplay of DNA replication and loop extrusion.

It has been showed that topoisomerase II β (TOP2B) interacts with cohesin and CTCF at TAD borders[151] and its binding is dependent on cohesin[152]. TOP2B was found to accumulate specifically after the CTCF barrier (*i.e.* just outside the TAD border). However, the exact role of topoisomerase at TAD border is not yet identified. TOP2B could be recruited at loop anchors to resolve loop extrusion-mediated knots or supercoiling, or travel together with the extrusion complex[153]. Replication and transcription processes both induce supercoiling and act as permeable barriers of loop extrusion. However, it is not yet clear to what extent the induced supercoiling or the very process of DNA replication or RNA transcription contribute to cohesin stalling.

Finally, other proteins have been shown to interact with cohesin (Yin Yang 1 (YY1), Mediator), but they were not found to stall cohesin on chromatin and act as barriers[88,104,154,155]. These proteins may help to reinforce cohesin-mediated interactions between specific genomic regions such as enhancers and promoters[104,154]. Because of the tension-dependent extruding activity of cohesin[113], densely protein-bound segments of chromatin (*e.g.* arrays of nucleosomes) may contribute to cohesin stalling, especially if their physical size is bigger than the cohesin step size (~40 nm)[104].

Thus, the loop extrusion model is more complicated than a simple dynamic extrusion halted by CTCF sites. Other proteins and nuclear processes can interact with cohesin and affect its extruding function to finetune genome architecture. To date, no completely impermeable barrier to cohesin was discovered, suggesting that cohesin is generally actively extruding loops rather than stalled on chromatin.

2.7 The interplay of loop extrusion and genome compartmentalization

Chromatin loops and compartments are two mesoscale structures of chromatin spanning a few hundreds of kb and a few megabases, respectively. These chromatin organization features are shaped by two different molecular mechanisms (**Figure 4**, **Figure 6B** and **Figure 8A**). However, since their length scale is roughly comparable, it is likely that their formation mechanisms interact with each other.

A/B compartments were observed to be strengthened in cohesin-depleted[31] or in Nipped-B-like (NIPBL, cohesin loader) depleted cells (**Figure 8B**)[156]. This argues for a 'mixing' role of loop extrusion, which counteracts genome compartmentalization. Moreover, after depleting cohesin from cells and allowing for cohesin resynthesis thereafter, the fastest loops to be resolved were located in the A compartment, while the slowest were found in the B compartment[31]. This finding was reinforced by the observation that silent chromatin states are less permissive to CTCF binding, likely due to a higher DNA methylation level[76]. Therefore, it is expected that loops in the A compartment are shorter and faster to form again than loops in B compartments, where extruding cohesins are less likely to be halted by CTCF[76]. This latter study also underlined the two-way interplay of loop extrusion and epigenetic mechanisms as both mechanisms can affect each other.

Therefore, cohesin-dependent loop extrusion counters the compartmentalization of chromosomes by allowing dynamic movements of chromatin across compartments and by 'mixing' chromatin. The complex relationship between these two processes still need to be clarified.

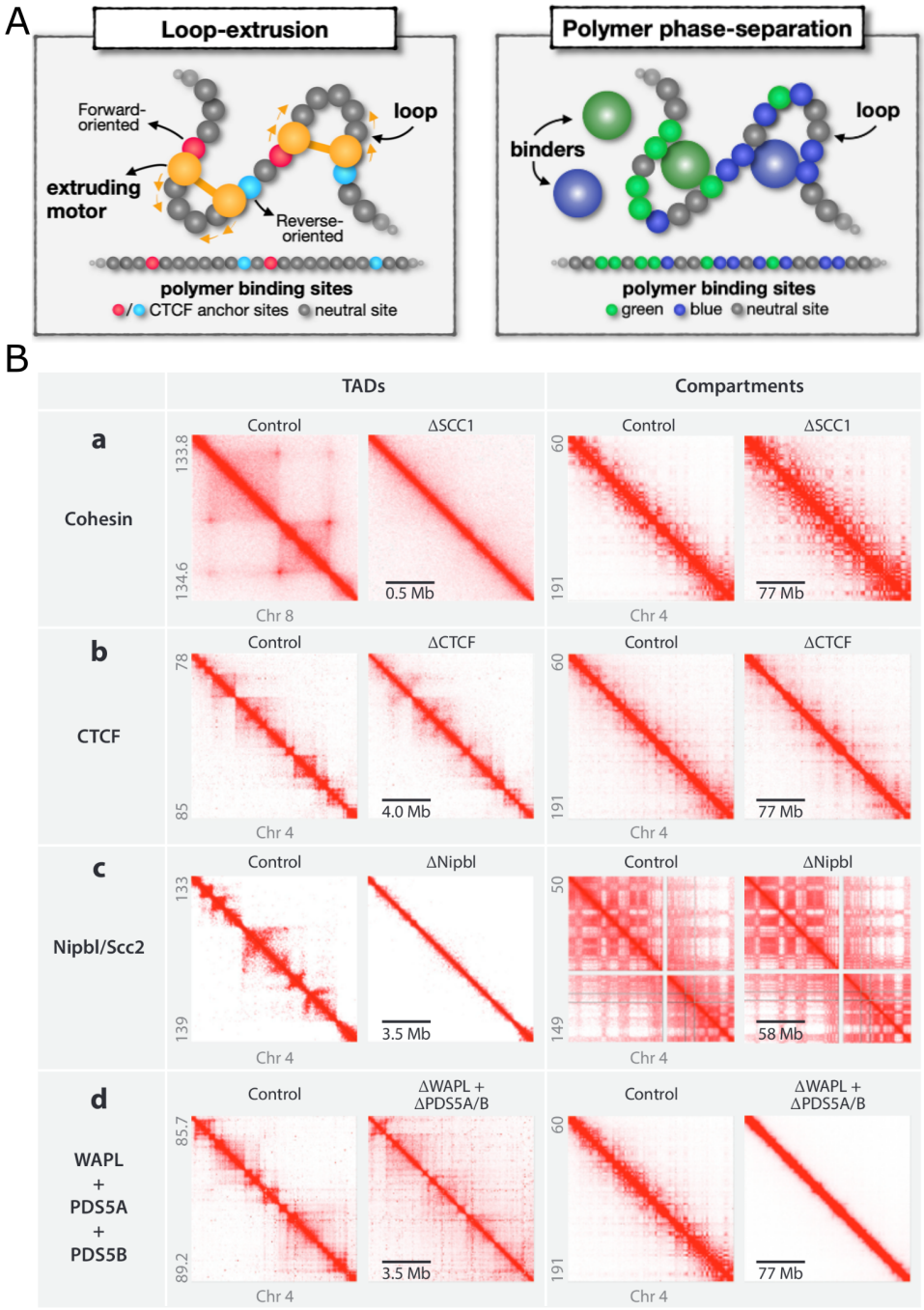


Figure 8: Genome organization is an interplay of loop extrusion and compartmentalization. A: Two different mechanisms (loop extrusion, left and compartmentalization, right) explain the folding of chromatin at the megabase scale. Cohesin extrudes loops by dynamically inducing interactions between monomers away from each other. Compartmentalization can be modelled by specific attraction potentials between beads of the same type (A or B compartment). Reproduced from [157]. **B:** Impact of depleting different proteins on Hi-C maps at the scale of TADs (left) or compartments (right). Depletion of RAD21 (a) or NIPBL (c) eliminates all loops and TADs from the genome, while compartmentalization is strengthened. Depletion of CTCF blurs TADs boundaries. Depletion of the cohesin unloader WAPL strengthens existing loops and creates longer loops, while decreasing compartmentalization. Reproduced from [69].

3. Loop extrusion dynamics is tightly regulated

The stochastic motion of chromatin, the finite processivity of cohesin and the permeable barriers contribute to chromatin loop dynamics. Cohesin chromatin binding kinetics and stalling rates are regulated by numerous molecules, in space and time.

3.1 Chromatin loops are likely to be dynamic rather than static structures

Several indirect evidence from various techniques suggested that chromatin loops emerging from loop extrusion are dynamic rather than static.

First, auxin-dependent depletion of RAD21 followed by an auxin washout showed that chromatin loops reappear quickly (about 1 hour) after cohesin synthesis (**Figure 5B**)[31]. This indicates that cohesin is able to quickly form loops throughout the whole genome.

Second, measuring CTCF and RAD21 residence times on chromatin by single-particle tracking showed that the two proteins have residence times orders of magnitude shorter than the typical 24 hours cell cycle duration (averages of 2 min for CTCF and 22 min for RAD21)[92]. Therefore, cohesin can go through multiple cycles of binding, extruding and unbinding events during a single G1 phase, thereby allowing loops to be extruded several times. The low CTCF residence time, as compared to cohesin residence time, can explain the nested TADs and loops found in Hi-C maps. CTCF sites may only be occupied by CTCF proteins 50% of the time on average[117]. Unbound CTCF sites would not block cohesin, which could continue extrusion and unbind or stop at the next encountered CTCF site, creating nested TADs at the population level (**Figure 6A**). Furthermore, absolute quantification of cohesin molecules in cells led to the conclusion that ~5.3 cohesin molecules / Mb are simultaneously bound on the genome[117,158], while the median loop size is ≈200 kb. Hence, it strengthens the possibility that loops are constantly extruded though cycles of cohesin binding and unbinding, and explains the high cell-to-cell heterogeneity observed in single-cell Hi-C[38] or in imaging[89,133,134,159,160].

Third, chromatin tracing of loops or TADs in single cells by DNA FISH demonstrated the high cell-to-cell heterogeneity in chromatin architecture[89,133–135,139,159,161]. These single-cell studies, along with single-cell Hi-C data[38,162], highlighted the fact that loops observed in bulk Hi-C maps are statistical averages while single cells display much higher variation in

chromatin architecture (**Figure 2C**). This observation underlines the role of stochasticity in genome organization, which possibly partly emanates from the dynamics of extrusion.

Fourth, polymer simulations best explained experimental Hi-C maps when dynamic rather than static loops were formed on the polymer[44,61,105]. Moreover, these simulations allowed to estimate the finite processivity and separation of cohesin complexes (both about 120 kb[32]), which were further confirmed by estimates from absolute quantifications of RAD21 molecules in single cells[117,158]. Consequently, loop extrusion dynamics could be partly controlled by modulating cohesin binding and unbinding rates on chromatin.

These converging results arising from various methods provided evidence for an understanding of chromatin loops as dynamic rather than static structures[19]. This dynamic nature of loops can be explained by the finetuning of cohesin binding kinetics and other molecules involved in loop formation at various timescales and by different molecules (DNA, RNA, proteins). During the course of my PhD, two different groups directly measured loop dynamics as function of time by fluorescently labeling loop anchors and found that loops were rare and transient[163,164]. These publications are discussed in section **5.4 First evidence of the dynamic nature of chromatin loops in living cells** and in the **Discussion**.

3.2 Cohesin binding kinetics is regulated by its protein partners

Cohesin binding is regulated at several timescales in the cell. First, the cohesin complex interacts with several proteins, which regulate the chromatin binding kinetics of cohesin during G1. Second, cohesin is involved in several distinct functions throughout the cell cycle. These different roles are tightly regulated by protein partners.

3.2.1 Regulation of cohesin binding and unbinding in G1

During the G1 phase, cohesin extrudes loops according to the loop extrusion model described above (**Figure 6B**). Various cohesin protein partners regulate its chromatin binding and unbinding rates, hence affecting chromatin organization.

NIPBL and MAU2 were identified as cohesin loaders, necessary for the loop extrusion activity of cohesin *in vitro* (**Figure 9A**)[106,113]. NIPBL depletion completely eliminated loops throughout the genome and copies a RAD21 depletion phenotype (**Figure 8B**)[156]. Although DNA sequence-encoded cohesin loading sites were not identified, the appearance of stripes

and jets in bulk averaged Hi-C maps of quiescent cells indicated that uni or bidirectionally extruding cohesins can be preferentially found at specific locations or extrude chromatin at different speeds on each side[91,102,103]. These specific locations harbored elevated NIPBL and H3K27ac ChIP-Seq peaks[102,103]. By controlling cohesin loading rate, NIPBL levels alter chromatin organization at the scale of TADs and compartments.

Cohesin variants co-exist in the cells depending on the STAG subunit integrated in the cohesin complex (**Figure 6C**). STAG1 or STAG2 were found to exhibit different contributions to cohesin-dependent genome topology. Cohesin-STAG1 variant was preferentially found at the TADs or loop boundaries and had longer residence times on chromatin, while cohesin-STAG2 was involved in cell-type specific transcription and was enriched at non-CTCF sites[165]. Therefore, these two cohesin variants have distinct dynamics, which might be the cause or the consequence of the different regions they are binding to and regulating[165]. Their differential incorporation in the cohesin complex regulates cohesin dynamics across the genome.

The cohesin unloading factor WAPL releases cohesin from chromatin. In accordance with the loop extrusion model, WAPL depletion leads to increased loop length and number of loops[95,166], as well as increased levels of chromatin-bound cohesin[95,166,167] due to higher cohesin processivity (**Figure 8B**)[61]. PDS5A and PDS5B proteins also act as unloading factors of cohesin and their depletion results in increased cohesin residence time[95,168,169] and increased long-range chromatin interactions[95], similar to a WAPL depletion phenotype. By increasing cohesin unbinding rates, WAPL and PDS5 proteins influence the length and insulation of chromatin loops (**Figure 9B**).

Finally, it was shown that CTCF and WAPL compete for the same binding site on cohesin (**Figure 9C**)[94]. Hence, the interaction between CTCF and cohesin could protect cohesin from WAPL-dependent unloading[99]. Despite the absence of direct evidence, a CTCF-cohesin interaction might stabilized the two proteins on chromatin, potentially prolonging the spatial interaction between anchors at TAD borders (**Figure 10E**).

These protein regulators finetune cohesin binding kinetics and therefore affect TAD and loop chromatin structures. This finetuning was well illustrated by co-depleting the cohesin loader NIPBL and the cohesin unloader WAPL, which have opposite effects on cohesin binding and unbinding rates. In this double depletion, the spatial overlap between

consecutive TADs was similar to the WT, while single depletion drastically affected this overlap (reduced for NIPBL and increased for WAPL)[170]. This experiment proved that loop extrusion dynamics needs to be correctly controlled to ensure proper spatial genome organization. The correct balance between cohesin binding and unbinding creates TADs and loops of physiological insulations and lengths and allows long-range chromatin interactions to occur at the required frequency.

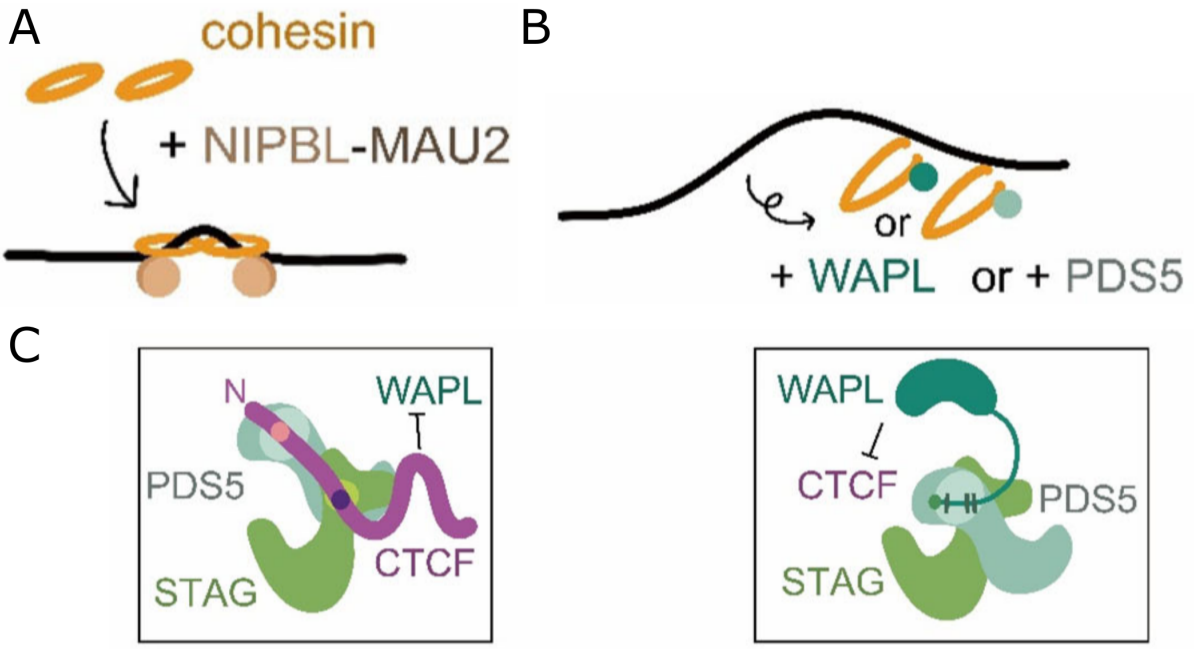


Figure 9: Protein regulators of cohesin processivity. A: NIPBL-MAU2 are cohesin loaders required for the loop extrusion activity of cohesin. **B:** WAPL and PDS5 proteins unload cohesin from chromatin. **C:** CTCF and WAPL compete for the same binding site on the STAG subunit of cohesin. Reproduced from [171].

3.2.2 Regulation of cohesin during replication and mitosis

Cohesin mediates sister chromatid cohesion (maintain the two chromatids together) between S phase and mitosis. Concomitantly to S phase entry, cohesin is acetylated by ESCO1 and ESCO2, which promotes sister chromatid cohesion[172,173]. This acetylation is dependent on ESCO1-PDS5 interaction[174] and protects cohesin from release by WAPL. This protection of cohesin from WAPL leads to increased cohesin residence time on chromatin after DNA replication, which allows efficient sister chromatid cohesion[175–177]. Also, STAG1-associated cohesins were found to be more acetylated and to have longer residence times than STAG2-associated cohesins, which would suggest that separated pools of cohesins exist

during the cell cycle (**Figure 10F**)[177,178]. Subsequent recruitment of sororin during G2 further stabilizes cohesin on chromatin[179]. Finally, upon entry in prophase, cohesin dissociates from chromatin due to sororin dissociation from chromatin by PLK1 phosphorylation[180,181]. Only centromeric regions remain bound by cohesin, which is released by separase in anaphase[180,181].

Although its exact role remains unclear, cohesin acetylation might also play a role in regulating cohesin residence time and loop length in G1[178,182].

The control of cohesin variants and cohesin processivity by protein partners enables cohesin to fulfill different functions (sister chromatid cohesion and loop extrusion) throughout the cell cycle. Thus, cohesin binding and unbinding rates are highly regulated at multiple timescales and cell cycle phases. Cohesin protein partners constitute a first layer of loop extrusion dynamics regulation, but other molecules contribute to this finetuned control of chromatin loop dynamics.

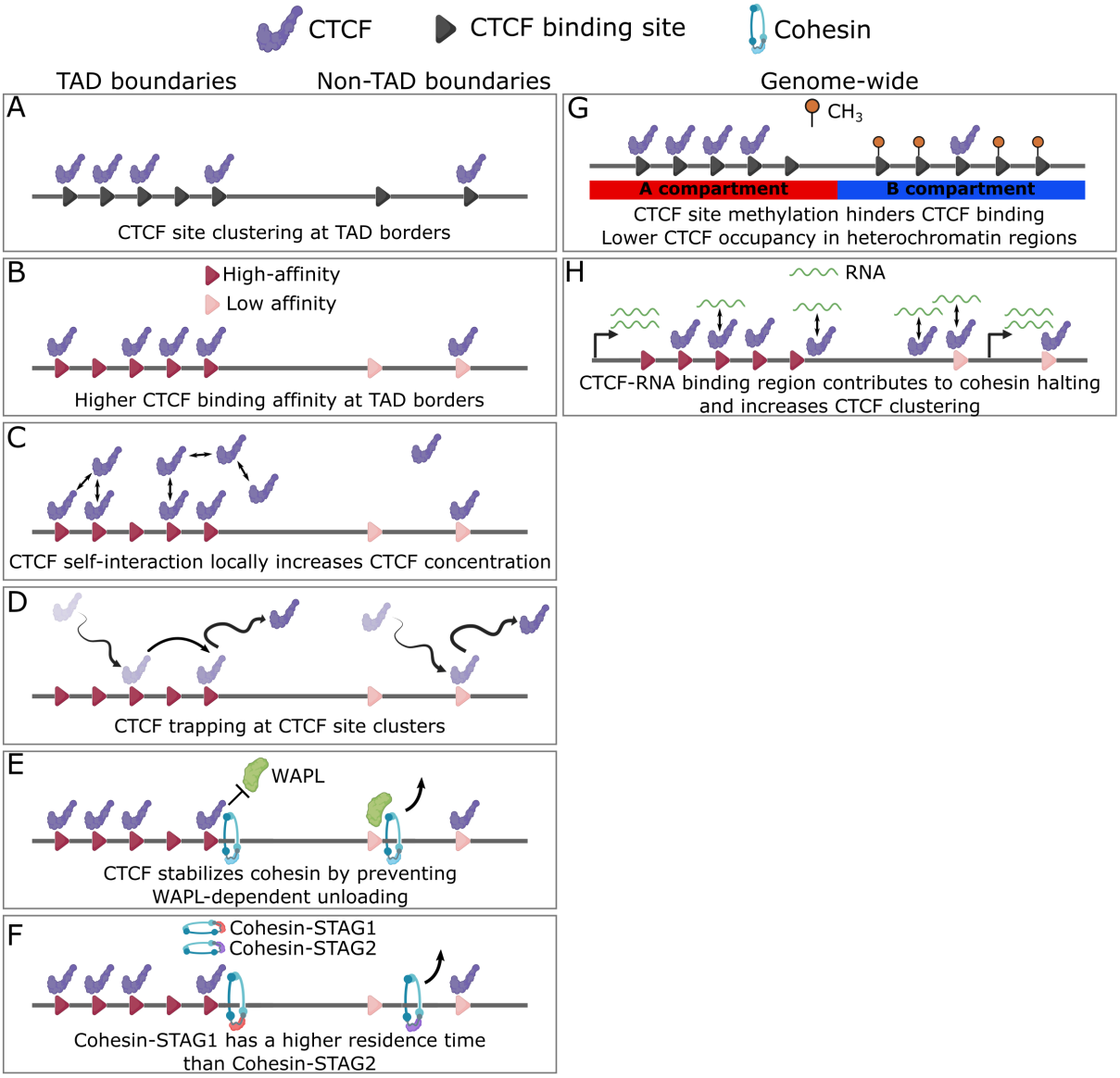


Figure 10: Mechanisms contributing to increased cohesin halting at TAD boundaries despite low CTCF residence time. A: CTCF site distribution is heterogenous genome-wide and CTCF site clusters are observed at TAD boundaries. **B:** CTCF sites at TAD borders display higher affinity to the CTCF protein. **C:** CTCF site clusters trap and alter CTCF diffusion, increasing its likelihood to bind on adjacent CTCF sites. **D:** CTCF can self-interact and can therefore locally increase its local concentration at CTCF site clusters, thus its likelihood to bind on nearby CTCF sites. **E:** CTCF and WAPL compete for the same cohesin binding site. In regions with CTCF clusters, cohesin has a higher probability of binding CTCF than WAPL. Its processivity is therefore increased at these clusters. **F:** Cohesin-STAG1 are more often found at TAD boundaries and exhibit higher residence times than Cohesin-STAG2 complexes. **G:** CTCF can bind RNAs and about half of the loops genome-wide are dependent on the CTCF RNA binding region. **H:** CTCF occupancy is lower in heterochromatin, perhaps due to CTCF site CpG methylation. Therefore, chromatin loops are longer and weaker in heterochromatin than in euchromatin.

3.3 DNA sequence-specific regulation of loop extrusion through CTCF binding strength and clustering

3.3.1 Differential CTCF binding strength: not all CTCF sites are equal

The CTCF protein binds to a 19 bp specific DNA motif (TGGCCACCAGGGGGCGCTA)[31,183]. However, DNA sequence motifs are averaged probabilistic sequences. Throughout the genome, some sequences will display the exact consensus motif, but the majority will display slight differences compared to the consensus motif. Therefore, CTCF might not bind with the same strength to different sequences. Moreover, because the CTCF binding motif is not palindromic, its orientation constitutes a code by itself to define loop anchor location.

By computing the binding affinity of CTCF binding sites genome-wide, CTCF sites located at TAD boundaries were found to exhibit a higher affinity and a higher CTCF ChIP enrichment than non-TAD boundary CTCF sites (**Figure 10B**)[184]. Therefore, the strength of CTCF binding is modulated by its DNA motif sequence and differs functionally throughout the genome. It was also shown that CTCF sites can initiate TAD boundaries outside of their genomic context[185,186]. Moreover, it is not only the CTCF site itself that can modulate CTCF binding affinity, but also its flanking sequences[185]. CTCF binding sites that are shared across different species, are more often found at TAD boundaries than at non-TAD boundaries, highlighting the importance of maintaining strong CTCF binding at TAD boundaries[184]. The affinity of CTCF binding for its binding motif could be converted into a cohesin permeability parameter. Indeed, the top 4000 CTCF ChIP-Seq peaks are more enriched in RAD21 and SMC3 peaks than the weakest 4000 peaks[32], showing that the binding affinity of CTCF regulates cohesin distribution or *vice versa*.

The differential binding affinity of CTCF on its binding sites also reflects functional differences. After incompletely depleting CTCF, some of the CTCF binding sites are lost, while others are retained[187,188]. The retained CTCF sites (high affinity) are mostly involved in architectural function (they are enriched at TAD boundaries and enriched at RAD21 CHIP-Seq peaks)[187,188], while the lost binding sites (low affinity) are located near Transcription Start Sites (TSS) and enhancers and might reflect a transcriptional role[188].

The CTCF motif contains CpG, which are subjected to methylation. Methylated regions of the genome are often associated with heterochromatin, and CTCF was shown to preferentially bind to open chromatin regions[76,189]. Methylated cytosines in CTCF binding sites decrease CTCF affinity[190,191]. It is not clear whether the methylation of the CTCF motif or the overall higher chromatin compaction is the cause of CTCF loss of occupancy in heterochromatin regions. However, heterochromatin regions can vary depending on cell types and therefore affect CTCF occupancy. Actually, it is possible to cluster cell types based on CTCF occupancy genome-wide, and tumor cells specifically display a higher CTCF site methylation probability than normal cells[191]. Therefore, CTCF occupancy and thus, chromatin loop location can be regulated through epigenetic modifications of the genome (**Figure 10G**).

This CTCF binding code is sufficient to predict 3D genome conformation as shown by the computational predictions of Hi-C maps based solely on DNA sequence[192,193]. In these models, the CTCF motif was identified as the main feature determining 3D genome organization.

The modulation of CTCF binding affinity by its DNA motif controls cohesin halting and the location of loop anchors throughout the genome. DNA binding sites locally regulate loop extrusion by preferentially halting or allowing cohesin to bypass CTCF sites[32]. The differential binding affinity, the orientation of CTCF sites and their methylation or location relative to AB compartments create a functional CTCF binding code to control chromatin loop borders. Finally, this CTCF binding code translates into structural and transcriptional roles of loop extrusion. However, how can CTCF sites halt cohesin when CTCF has a low residence time (2 min vs 20 min for cohesin[92]) on chromatin and is a permeable barrier?

3.3.2 CTCF clustering provides robustness to loop anchors

The DNA sequence distribution of CTCF binding sites, as well as the distribution of CTCF ChIP-seq peaks is not homogeneous across the genome. Some regions of the genome display high CTCF site clustering and accumulate CTCF binding, while others only display sparse single CTCF sites and are depleted from CTCF. CTCF clustering could compensate for CTCF permeability and low residence time and provides robustness to loop anchors.

TAD boundaries generally do not consist in a single pair of convergent CTCF sites but rather in clusters of multiple CTCF binding sites that cooperate redundantly to provide

boundary robustness (**Figure 10A**)[100,184,194]. Moreover, clusters of CTCF sites co-localize more often with RAD21 than single CTCF sites, independent of their relative position to TAD boundaries[184]. Also, single CTCF sites are more often found close to TSS than clusters, which could explain why low-affinity CTCF binding is associated with transcriptional binding, while high-affinity CTCF binding is associated with architectural functions[184,188,194] (**Figure 10A**). Finally, these CTCF clusters are populated by more CTCF sites when they are conserved across species[184,195], underlining their importance in maintaining chromatin architecture.

Super-resolution imaging provided estimations of CTCF and cohesin clustering in cells. CTCF protein clusters contained approximately 2-8 molecules, while cohesin formed clusters of about 3-15 molecules. In addition, 25% of CTCF clusters were coupled with cohesin clusters at distances of about 65 nm, similar to the cohesin ring size (50 nm)[92,93]. This clustering could be mediated by the self-interaction ability of CTCF (**Figure 10C**)[147]. Despite the difficulty to image CTCF sites and the CTCF protein simultaneously, we can speculate that the CTCF protein clusters are located on CTCF site clusters at TAD borders, suggesting that several cohesin and CTCF molecules maintain the TAD borders in close proximity.

Moreover, more than 60% of CTCF ChIP-Seq peaks overlap multiple CTCF binding sites[194], suggesting that CTCF site clusters attract multiple CTCF proteins. Because CTCF has a short residence time on chromatin (2 min on average)[92], the probability that an extruding cohesin encounters a bound CTCF motif is low if the boundary is marked by a single CTCF site. However, clustering of CTCF sites increases the probability for cohesin to encounter a bound CTCF site (**Figure 10A-C**). Furthermore, several CTCF proteins could bind multiple CTCF sites of a cluster at the same time, consequently increasing the duration of cohesin blocking at this location. Another role of CTCF site clustering might be the spatial trapping of CTCF[196]. CTCF site clusters attract and trap CTCF by offering multiple binding sites in a small space. This alters CTCF diffusion and locally increases its concentration (simultaneously increasing the probability of CTCF to self-interact[147]) and likelihood to bind (**Figure 10D**)[196].

The non-homogeneous distribution of CTCF binding sites and their clustering enable them to finetune chromatin loop dynamics by halting cohesin with different probabilities, depending on their size and binding affinity. Also, CTCF clusters increase the probability of recruiting CTCF locally by hindering its diffusion, therefore increasing its local concentration

by CTCF self-interactions. Therefore CTCF site clusters potentiate CTCF ability to halt loop extrusion despite its short chromatin residence time. Thus, sequence-encoded information instructs loop extrusion to shape chromatin architecture through CTCF binding site affinity and clustering.

3.4 RNA-dependent regulation of loop extrusion

RNA Pol II was proved to be a mobile barrier of loop extrusion. However, the RNA molecule itself can affect chromatin loop dynamics.

CTCF can bind RNA with an affinity at least two orders of magnitude higher than DNA[197]. Its RNA interactome is not specific but CTCF can bind thousands of transcripts with high affinity ($K_d \approx 0.5$ nM)[197]. Specific zinc fingers (ZF1 and ZF10-11) of the CTCF protein mediate its RNA-binding ability. The interaction between CTCF and RNA is involved in CTCF binding to DNA since deletion of the RNA-binding domains in CTCF disturbs its chromatin binding[147,198], its self-interaction and clustering ability[147]. Furthermore, due to alteration of chromatin binding, TAD organization is modified upon deletion of the CTCF RNA-binding regions. A significant fraction of chromatin loops (about 45%) disappears when CTCF cannot interact with RNA, and most chromatin loops become weaker, even if CTCF and cohesin are still bound to chromatin anchors (**Figure 10H**)[147,198].

More specifically, long non-coding RNAs (lncRNAs) were identified to interact with the loop extrusion machinery[199]. The *HOTTIP* lncRNA was found to reinforce CTCF boundary and maintain TAD structure by mediating R-loops (DNA-RNA hybrids) and its targeted degradation at the boundary impaired TAD structure[199]. An *in vitro* study has identified that R-loops slow cohesin translocation, which could partly explain how non-coding RNAs affect loop extrusion[200]. Enhancer RNAs (eRNAs) are produced by enhancer transcription. In a specific case, an eRNA was found to interact with cohesin, as well as to recruit and maintain cohesin at the Myogenin locus, although the interaction with CTCF was not assessed in this study[201]. More studies are needed to better understand how non-coding RNAs affect chromatin loops and TAD structure. Nonetheless, these initial results open new possibilities in the regulation of chromatin organization.

Thus, in addition to RNA Pol II, RNA molecules are able to alter chromatin loops and TADs, mostly through their interaction with the RNA-binding domains of CTCF.

The extrusion process appears to be simple: cohesin is the motor of loop extrusion and extrudes loops until encountering an obstacle or unbinding from chromatin. However, the regulation of cohesin dynamics is tightly controlled and complex (**Figure 9** and **Figure 10**). Its binding kinetics is regulated by several partner proteins that can either shorten or extend its lifetime on chromatin, therefore controlling loop length, loop insulation and frequency of extrusion. CTCF, among other molecules, is a barrier to cohesin and defines the boundaries of TADs. The short residence time of CTCF on chromatin and its small size seem to argue that CTCF is a poor barrier molecule. However, CTCF affinity to its binding motif, CTCF site orientation, CTCF site clustering which increases CTCF local concentration and self-interaction and CTCF interactions with RNAs all can potentiate CTCF's ability to halt cohesin. Thus, chromatin loop length and insulation, as well as frequency of extrusion are tightly controlled through interactions with numerous molecules in the nucleus. Moreover, this tight control of chromatin loop dynamics is conserved across species[184,202–205]. The conservation of this mechanism indicates that chromatin loop extrusion plays key functional roles in the nucleus[202].

4. Loop extrusion as a process to adjust long-range chromatin interactions in various nuclear functions

4.1 Cohesin and its regulators are mutated in diseases

The evolutionary conserved mechanism of loop extrusion as well as its conserved regulation mechanisms underlines its functional importance. Mutations of the cohesin complex lead to severe genetic developmental diseases such as Cornelia de Lange syndrome and Mungan syndrome[206]. These diseases are commonly termed cohesinopathies and exhibit major phenotypes such as mental retardation, facial dysmorphism, among other symptoms[206]. These observations highlight the major role of cohesin and potentially loop extrusion in cell physiology.

Almost all members of the cohesin complex, as well as its protein partners, were found to be significantly and recurrently mutated among 33 cancer types. Namely, STAG2, SMC1A, SMC3, RAD21, CTCF and NIPBL were identified as cancer driver genes[207,208]. Moreover, CTCF binding is altered in cancer, as more CTCF peaks are found genome-wide in cancer cells than in primary cells[129,209]. Although the specific cohesin function(s) promoting cancer were not identified in these studies, these results highlight the essential role of cohesin and its partners in the physiological integrity of the cell.

Since cohesin is involved in multiple nuclear functions, the outcome of its mutation is complex to study. Understanding the multiple functional roles of cohesin and loop extrusion in the nucleus will allow us to better understand its dysfunctions in diseases. Specifically, cohesin partners involved in the regulation of loop extrusion dynamics such as STAG2 or NIPBL were also found to be mutated in diseases. Therefore, characterizing the dynamics of loop extrusion might shed light on potential mechanism(s) of nuclear dysfunction leading to abnormal phenotypes.

4.2 Loop extrusion can help to regulate gene expression

One of the main hypothesized roles of loop extrusion is to regulate gene expression. For instance, genomically distant enhancer and promoters could be brought into close spatial proximity by loop extrusion, hence activating transcription (**Figure 11A**). Conversely, the relative insulation of TADs created by loop extrusion could prevent ectopic gene transcription

by keeping enhancers and promoters that should not get into contact away from each other (**Figure 11A**). Although, this simple view of loop extrusion-dependent gene regulation is challenged by experimental observations, some evidence argues in favor of this model. Below, I describe the different findings highlighting the context-dependency of gene regulation mediated by loop extrusion.

4.2.1 Enhancer-promoter interactions are favored within TADs and prevented between TADs

Enhancers are short DNA regions that can be bound by protein activators to activate gene transcription. The exact mechanism of how enhancers can activate the transcription of genes located hundreds to thousands of kb away from the enhancer location is not well defined. The increase of long-range chromatin interaction frequencies mediated by loop extrusion might be a useful process to allow enhancers to find and regulate their target promoter, or on the contrary impede ectopic activations by insulating promoters and/or enhancers into different domains (**Figure 11A**)[210].

It is noteworthy that Enhancer-Promoter (EP) interactions are able to form loops by themselves without cohesin or CTCF[27,88,211], and RNAs are also able to shape long-range DNA contacts[212]. Genome-wide analysis of Micro-C data showed that although cohesin-mediated loops are the strongest loops detected (for long-range interactions), loops can also form through EP, promoter-promoter (PP) interactions or at TSS[27,88]. These EP and PP loops are slightly smaller (100 kb) than cohesin-mediated loops (160 kb)[88]. Moreover, the strength of EP, PP and TSS loops correlates with gene expression levels and these loops are abrogated by RNA Pol II inhibition, while cohesin-dependent loops remain unchanged[27,88]. Although these loops represent a significant part of the loops detected in Micro-C maps[88], their description is beyond the scope of this introduction, and I will focus solely on cohesin-mediated gene regulation.

Using genome-wide studies, it was shown that genes dysregulated by cohesin depletion were more often found near TAD boundaries and, among these genes, their fold-change correlated positively with their spatial proximity to a TAD boundary[139]. Moreover, genes located in the same TAD shared a higher correlation in their transcriptional dynamics than genes in different or random domains[87]. Cohesin depletion also led to dysregulations of inducible genes rather than housekeeping genes, during differentiation or external stimuli[213–217]. Also, viral infections triggered modifications of chromatin architecture and loop extrusion

alterations to easily replicate[218–220]. Moreover, despite CTCF sites being present on the DNA sequence of all cells in a single organism, differential chromatin looping was observed between cell types[167,217,221–224], and these differences could be effectively used to cluster cell types[217]. Different cell types can harbor various epigenetic modifications on the same chromatin region, likely creating differences in CTCF binding. Differential binding of CTCF depending on chromatin state[27,76] might explain partially the transcriptional differences in chromatin loop location and insulation across cell types. Therefore, loop extrusion may play a role in cell type-specific transcriptional programs. These studies suggested that loop extrusion is mainly required for changes in transcriptional programs rather than steady-state transcription and is important for tissue-specific gene expression.

Using synthetic gene activation by targeted recruitment of a transcription factor, it was shown that cohesin was required to activate gene transcription only in conditions where the enhancer was located away (at least 400 kb) from its target promoter. In the same study, depletion of CTCF did not trigger transcriptional changes. This suggests that the extrusion process itself, rather than the integrity of border elements, drove the observed transcriptional changes, possibly by decreasing EP distance[225]. Strengthening these results, several studies found that genes located in loops of bigger sizes are more sensitive to cohesin depletion[211,213,226], suggesting that cohesin is only required for long-range, rather than short-range, control of transcription.

CTCF sites can act as insulators when located between an enhancer and its target promoter, and the disruption of loop extrusion leads to ectopic transcription in these specific cases (**Figure 11A**)[227–230]. Indeed, deletions and inversions of CTCF sites in TADs lead to severe developmental phenotypes through ectopic enhancer-driven gene transcription[231,232]. Chromatin loop extrusion can therefore insulate enhancers from promoters to prevent ectopic gene expression.

By systematically displacing an ectopic enhancer from its promoter at various genomic distances within a TAD, the expression level was found to heavily depend on EP distance (**Figure 11B**)[211,230] and followed a sigmoidal law with respect to contact frequencies (**Figure 11C**)[230]. The sole addition of the enhancer initiated a local self-interacting domain. But the addition of either a triplet of strong CTCF sites, or the presence of cohesin, creating a smaller domain encompassing the enhancer and the promoter, drastically increased both contact frequencies within the created domain, as well as gene expression level[211]. As already

described, loop extrusion was only required for long-range gene activation and was dispensable for short-range enhancer action[211]. Conversely, insulation of the enhancer by inserting a CTCF boundary greatly reduced the transcriptional output[185,230]. More strikingly, this decrease in gene expression was proportional to the number of CTCF inserted to create the boundary[185]. Moreover, in the cell lines where the enhancer was mobilized outside the TAD, no transcription was detected, as if the TAD was sufficient to prevent transcription from the enhancer[230]. Finally, by placing two different promoters and a single enhancer within a domain, it was possible to assess how enhancers regulate multiple promoters. In this configuration, a smaller domain size helped the competing promoter to be transcribed, showing that domain size can help to target enhancers to their promoters[211]. Using synthetic constructions, these studies showed a clear dependence of the transcriptional output on the enhancer-promoter contact frequency, and the ability of loop extrusion to act as an insulator or to foster EP interactions.

Thus, by insulating or favoring EP interactions, chromatin loop extrusion can modulate gene expression.

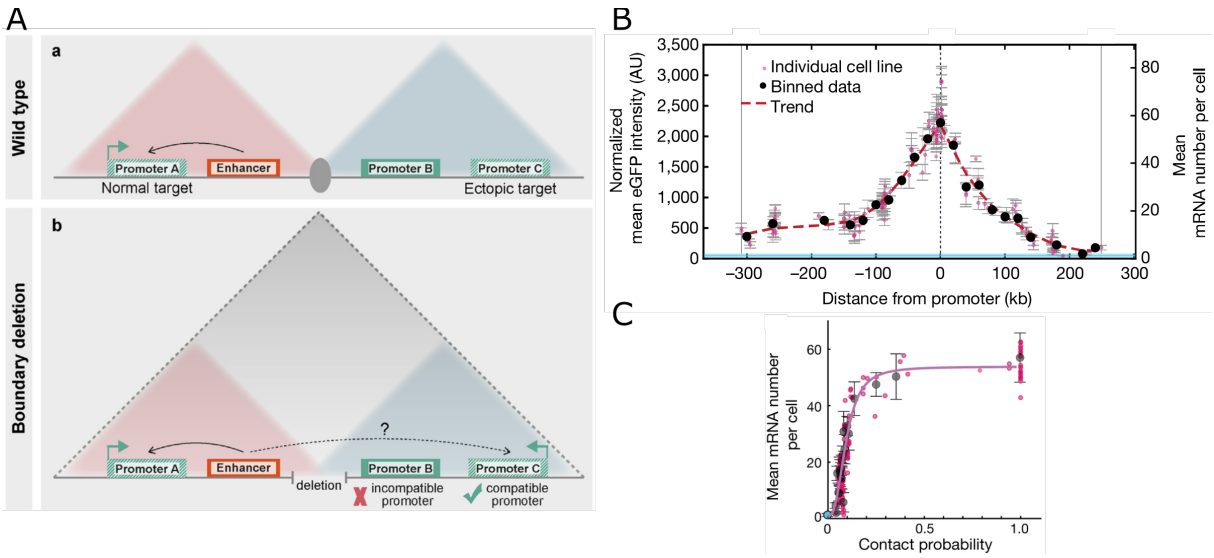


Figure 11: Loop extrusion may control gene expression through enhancer-promoter interactions. A: Simple model of gene expression control by loop extrusion. Cohesin-dependent TADs constrain enhancer-promoter contacts and only enhancer and promoters located within the same TAD can come into contact (top). Removing the TAD boundary merges the two TADs together and expression from gene C can occur (bottom). However, further investigation showed that this model was not universal and more complex (see below). Reproduced from [233]. **B:** An ectopic enhancer was systematically moved within a TAD. The level of expression correlates negatively with distance to promoter. **C:** Mean mRNA number

as function of contact probability. A sigmoid function (best fit) is plotted in light red. Panels **C** and **B** are reproduced from [230].

4.2.2 Enhancer-promoter contacts and gene expression are robust to changes in genome organization

Altering the loop extrusion process and its insulation ability can lead to transcriptional changes. However, several lines of evidence argue against these observations and show that gene expression can be robust to chromatin organization changes.

The global deletion of CTCF or RAD21 in steady-state cell lines only leads to modest changes in nascent gene transcription[31,88]. In these cell lines, gene expression is assessed a few hours after protein depletion, as compared to days or many cell generations in developmental studies or in studies using genetic modifications such as CTCF deletions. It was shown that transcriptional levels were not stabilized more than a month after chromatin topology perturbation in mouse embryonic stem cells (mESCs)[211]. Despite modest changes in average gene expression, the variance of gene expression across single cells was higher when cohesin was depleted[140], arguing in favor of a model where chromatin architecture finetunes gene expression levels. The temporal dependence of transcriptional changes might explain the discrepancies in interpreting the relationship between genome organization and gene regulation.

The depletion or inversion of CTCF sites during development was assessed in the Hox gene cluster, a cluster of genes whose order on the chromosome is the same as their spatial expression in the developing embryo. This spatial collinearity in the expression of Hox genes was maintained upon CTCF site deletion or inversion, despite abnormal gene expression boundaries. However, the timing of expression was altered by these TAD modifications[234,235]. Although TAD insulation was disrupted in the mutated embryos, enhancers could bypass the TAD border and promote proper gene expression[234–237]. Other developmental studies revealed that deletions of CTCF sites resulted in TAD fusions, as expected. However, these genome topology perturbations only led to moderate changes in gene expression and spatial patterns of gene expression[238–241]. Hence, TAD insulation might enable to finetune the spatio-temporal expression of genes by providing robustness and precision rather than being essential for turning transcription on and off.

The model in which enhancers need to be in close spatial proximity with the promoter to activate gene transcription was also contradicted by several lines of evidence. The disruption of loop extrusion by CTCF site deletions or mutations showed that changes in EP spatial distance generally led to only modest changes in gene expression[240,242–244]. Also, no correlation was found between the EP spatial distance and the transcriptional level of the Sox2 endogenous locus in living mESCs[244].

Thus, gene expression and EP interactions can be robust to chromatin structure changes and bypass CTCF borders to promote gene expression, albeit with reduced precision.

4.2.3 Reconciling experimental observations

Despite the apparent discrepancies between experimental observations in different model systems and at various timescales, a recent attempt to reconcile the different findings that genome folding can or cannot regulate gene expression has been made[138]. I summarize below its main findings.

As already mentioned, the genomic context of the studied TAD is important. CTCF sites are found across the whole genome and some of them correspond to TAD borders, while others do not (**Figure 10**). Using endogenous CTCF sites in a reporter assay, it was possible to show that not all CTCF sites can insulate or promote gene expression and this ability depended on their genomic context[185,186]. Therefore, loop extrusion control over gene expression is context-dependent, and trying to define general rules might be difficult. One paradox is that the insulation of TADs created by loop extrusion is only about 2-fold (**Figure 12A**), while changes in transcriptional levels can be 10-fold or more upon gene activation[138]. How can a subtle change in chromatin structure create such large changes in transcription[138]?

The observation that depletion of cohesin or CTCF only mildly alters gene expression was made after a few hours of depletion (**Figure 12B**)[31,88]. But developmental assays or CTCF deletions that assayed transcription after days or many cell divisions often display higher transcriptional changes (**Figure 12B**)[228,230,245]. Adding complexity in the temporal relation between chromatin topology and transcription, correlations between enhancer-promoter proximity and transcriptional output might not be observed despite the two being dependent on each other. Indeed, it is possible that the frequency and temporal accumulation of transient contacts rather than prolonged enhancer-promoter contact are the main regulator

of transcription activation. In order to observe this, a large amount of single-cell temporal data needs to be acquired[138]. Functional processes, such as transcription, that depend on chromatin architecture might exhibit hysteresis (*i.e.* the system has memory and its behavior depends not only on its current state, but also on previous states). A temporal delay between causes (chromatin architecture changes) and consequences (transcriptional changes) might occur, hence obscuring experimental correlations between chromatin architecture and transcriptional output.

Using modeling, it is possible to explain this paradox. Experiments showed that the transcriptional output has a sigmoidal relationship with contact frequency (**Figure 11C**)[230]. Depending on the sigmoidal regime (super-linear or sub-linear) of the studied locus, small changes in TAD structures could either trigger a massive transcriptional response (super-linear regime) or no response at all (sub-linear regime) (**Figure 12C**)[138]. The two outcomes are still in agreement with a model where promoters activate transcription in a contact-dependent manner, possibly explaining the different and seemingly conflictual experimental observations. Indeed, certain 3D topology perturbations might show no effect on transcription because the promoter is in the sub-linear regime. This model also exhibited hysteresis: the transcription rate and the location of the sub- and super-linear regimes were modified depending on the initial state of transcription (gene already transcribing or gene not expressed at the start of genome architecture changes)[138]. Hysteresis might explain the apparent modest changes in gene expression after global disruption of chromatin architecture on the timescales of hours (*e.g.* RAD21 or NIPBL depletion[31,156]) (**Figure 8B**). In these conditions, genes may retain their initial transcription state after a few hours, but such memory may be lost after days. We could speculate that the memory may be provided by epigenetic changes for example, which would be compatible with the timescales considered.

Despite the seemingly contradictory results about the role of loop extrusion in gene expression, it may be possible to unify them in a model where loop extrusion regulates gene regulation in a context-dependent manner[138], but is not the only regulator of EP interactions. Tight genome folding regulation can enable precise and robust spatio-temporal gene expression by controlling the possible interactions between enhancers and promoters. The influence of loop extrusion on gene regulation is still an active domain of research and many mechanisms are yet to be understood.

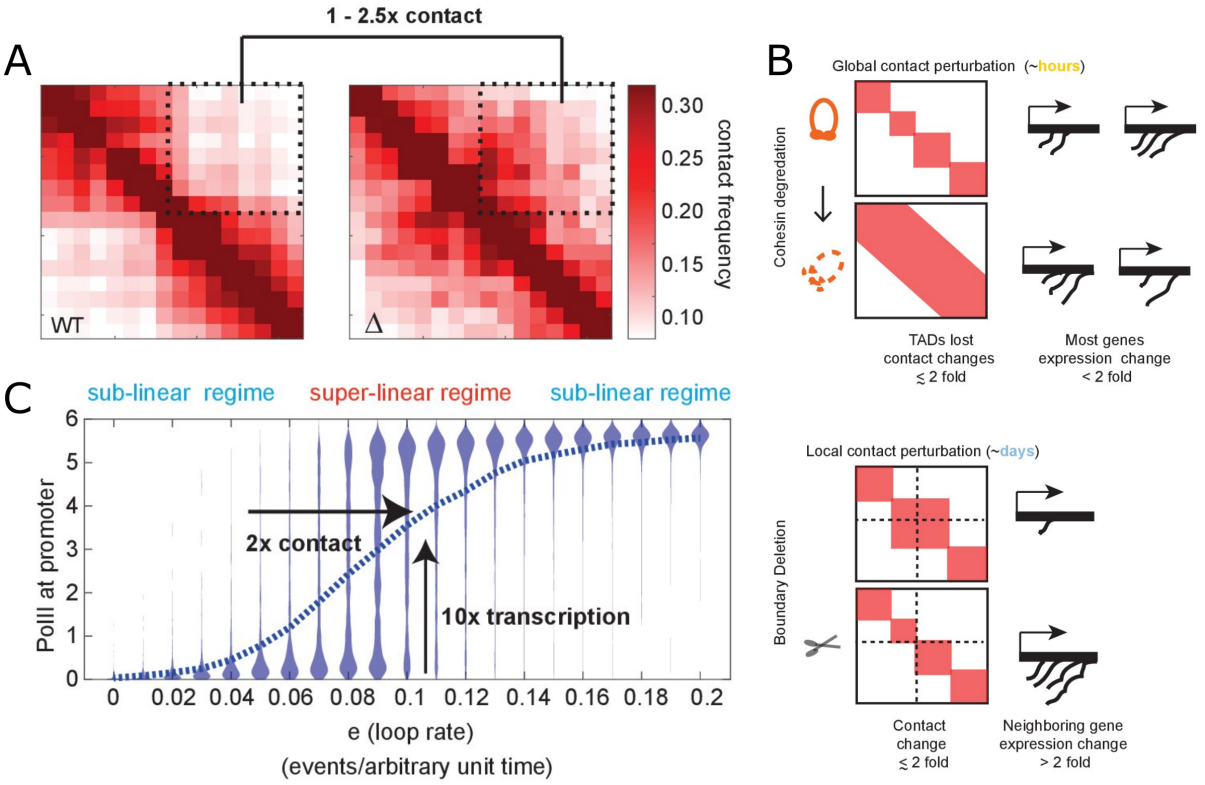


Figure 12: Subtle changes in contact frequency can lead to large changes in transcriptional levels. A: Hi-C maps of *Drosophila* embryos at Hox genes in WT embryos (left) and in embryos where the domain boundary was deleted (right). Only 1-2.5 fold change in contact frequency was observed between the two maps. **B:** Seemingly contradictory observations that a global disruption of chromatin architecture only leads to modest changes in gene expression (top), while local disruption of chromatin architecture over several days can lead to large changes in transcription (bottom). **C:** These seemingly contradictory observations (**B**) can be explained by a model exhibiting a sigmoidal function between transcription output (PolII at promoter is used as a proxy) and contact frequency (loop rate is used as a proxy). In the super-linear regime, small changes in contact frequency can lead to large changes in transcription. These results are from a computational model. All panels are reproduced from [138].

4.3 Loop extrusion occurs at DNA repair foci

Recent evidence that chromosome organization and cohesin-mediated loop extrusion are involved in DNA repair extended the functional roles of this process[196,246–248].

One of the first steps to repair damaged DNA is the deposition of γH2AX on chromatin by phosphorylation of the histone variant H2AX[249], which forms foci in the damaged nucleus (**Figure 13A**)[250]. The deposition of these histone marks was found to be restricted to TADs and stopped at their boundaries[247]. These marks are deposited over megabase-scale in dozens of minutes in a cohesin-dependent manner, while the DNA repair

machinery is recruited and anchored at the damaged DNA (**Figure 13B**). Stripes emanating from the double-strand break (DSB) were visible in Hi-C maps as expected from a unidirectional scanning of chromatin[247]. It is not clear whether cohesin is the motor of the DSB-dependent loop extrusion and deposition of γH2AX marks since knockdown of a cohesin subunit only slightly decreased γH2AX deposition profiles[247]. Other loop extruding complexes (*e.g.* SMC5/6[251]) could also help or be the main motors of extrusion at DSB sites.

At DSB sites, cohesin could restrain chromatin movement and therefore limit the DNA target search to neighboring DNA. This movement restriction might prevent aberrant repair with false homologous pairing by promoting *cis* interactions[251,252]. Polymer simulations showed that targeted loading and stabilization of cohesin at DSB sites would allow efficient DNA repair as experimentally observed (**Figure 13C**)[253]. A recent study where ESCO2 was found to be recruited at DSB sites and acetylated, hence stabilizing, SMC3 consolidated these theoretical considerations[248].

Some recent evidence describes loop extrusion as a mechanism to help repair DNA breaks. However, more work is required to better understand to what extent loop extrusion is involved in DNA repair and the exact molecular mechanism of the interplay between the two processes.

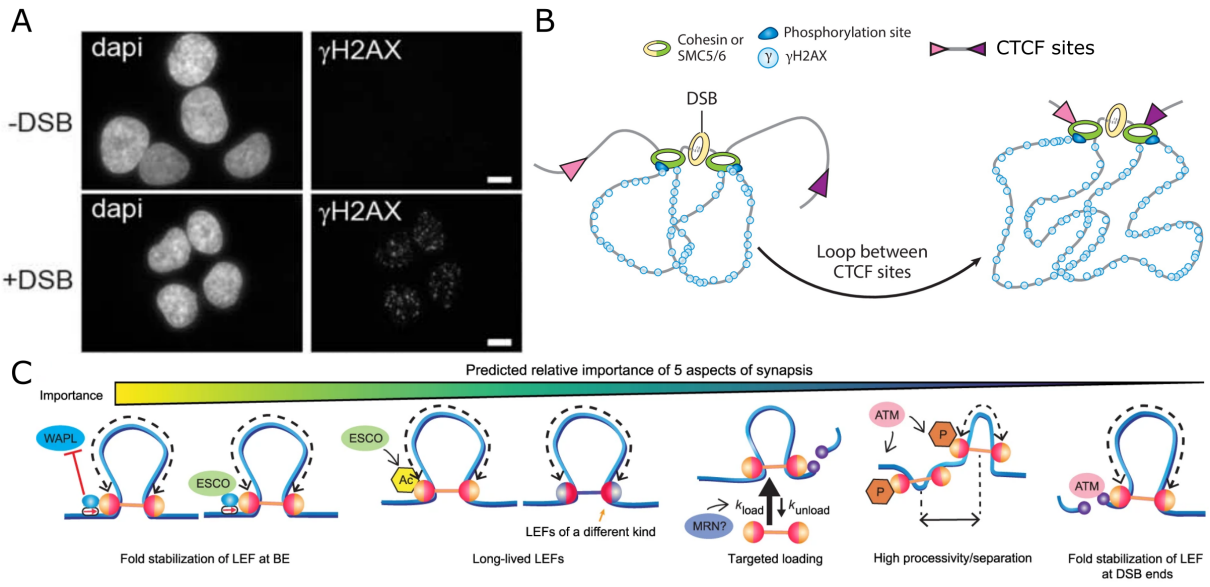


Figure 13: Loop extrusion helps repairing DNA damage. A: Upon DNA damage (+DSB) H2AX foci appear in the cell nucleus. Reproduced from [247]. **B:** Model in which loop extrusion helps to phosphorylate the H2AX histone variant within a damaged TAD. The deposition of γH2AX marks stops at TAD boundaries. Adapted from [104]. **C:** Theoretical possible

mechanisms of cohesin dynamics changes upon DNA repair, inferred from polymer simulations. Reproduced from [253].

4.4 Maximal diversity of antibodies is ensured by loop extrusion during V(D)J recombination

B cells are among the few cells that rearrange their genome during their lifetime. V(D)J recombination allows them to produce an almost infinite number of unique DNA sequences, which translates into a unique antigen for each B cell clone. This large diversity of antigens is obtained by two mechanisms: V(D)J recombination and somatic hypermutation[254]. Different segments of the genome can stochastically recombine to form the antigen sequence. Moreover, some of these sequences are hypermutated on purpose to create randomness in the resulting antigen sequence.

Loop extrusion plays a fundamental role in the recombination process of V and DJ segments. While dozens of V sequences can be recombined with DJ segments, a single V sequence will be part of the final antigen sequence. In the genome, V sequences form an array and they are separated by convergently oriented CTCF sites. Cohesin can scan this array and depending on CTCF site binding at this moment, the recombination will stochastically occur in various segments of the V segments array[255,256]. The ability of loop extrusion to scan sequences separated by long genomic distances is used in V(D)J recombination to ensure maximal diversity of generated antigens. This process highlights the importance of loop extrusion dynamics since the binding kinetics of CTCF controls the sequences used in V(D)J recombination.

Apart from these processes occurring in G1 phase of the cell cycle, chromatin loop extrusion and extruding SMC complexes are involved in sister chromatid cohesion, the firing and positioning of replication origins and the organization of mitotic chromosomes[98,104,257].

The process of loop extrusion has the ability to increase long-range chromatin interactions. Several described nuclear processes exploit this ability to allow contacts between sequences that are unlikely to come into close proximity by random collisions. But other processes, that are still to be discovered, might also take advantage of loop extrusion to favor

long-range contacts. The dynamics of loop extrusion seems to be more important than the actual formation of stable loops for these processes. Therefore, the quantification of chromatin loop extrusion dynamics is key to understand how loop extrusion helps achieving genome integrity, expression and replication.

5. Visualizing chromatin loop extrusion dynamics

5.1 Limitations of fixed cell techniques

The study of genome organization essentially relied on fixed cells (DNA FISH) and bulk-averaged techniques (Hi-C, ChIP-Seq)[69]. These methods generally use cell fixation. The effect of cell fixation and its potential induction of artifacts is yet understudied, while of major importance to understand how it affects experimental data. It was found that fixation can alter phase separated compartments[258], as well as contact probability curves in Hi-C data[34]. Furthermore, albeit powerful, these techniques lack temporal resolution, although time is a key component to understand loop extrusion function.

For instance, CTCF ChIP-Seq tracks generally exhibit CTCF binding at most identified CTCF binding sites. However, at a single timepoint in a single cell, it is likely that CTCF is only found at a fraction of its binding sites[92]. The low CTCF residence time and the low probability of a CTCF site to be bound explain how cohesin can bypass CTCF site barriers. CTCF residence time and binding probability represent two of the numerous regulations of loop anchor location and insulation. However, this information is hidden in averaged genomic assays.

As indicated above, there is evidence that the dynamics of extrusion itself is exploited by biological processes rather than the final loop structure. Hence, it is necessary to study chromatin loop extrusion in living rather than fixed cells as the temporal study of this process is of major importance to understand its functions.

5.2 Chromatin loop extrusion was visualized and quantified *in vitro*

Knowing that cohesin was the motor of loop extrusion, it was possible to determine the minimal components required for cohesin-mediated loop extrusion. Using purified proteins, *in vitro* DNA loop extrusion was made possible by the sole addition of cohesin, NIPBL and ATP to DNA[106,113]. This showed that extrusion was an active process that needs energy and does not act by a molecular ratchet fueled exclusively by Brownian collisions[106,113].

These assays revealed fundamental mechanisms of the extrusion process described above (*e.g.* bidirectional extrusion, monomeric and oligomeric extrusion (**Figure 7**)). Also,

they allowed to estimate the loop extrusion rate to be between 0.5-1 kb/s, demonstrating that cohesin is a fast motor (*e.g.* as compared to the slower 1 kb/min of RNA Pol II for instance)[106,113].

Although of major importance for the understanding of chromatin loop extrusion mechanism, these assays were performed *in vitro* with purified proteins and short segments of DNA. It is difficult to extend these results to living cells, where the crowded chromatin environment, as well as the numerous cohesin partners could potentially decrease or increase loop extrusion speed by an unknown factor. Consequently, fully describing chromatin loop extrusion requires a quantification of this process directly in living cells.

5.3 Visualizing specific chromatin loci in living cells

Quantifying chromatin loop extrusion in living cells requires to track and follow specific chromatin loci as function of time. Fluorescently labelling chromatin while keeping cells alive requires the use of a non-invasive method of visualization.

5.3.1 Usual methods to stain DNA are lethal for cells

Imaging specific DNA loci in living cells is challenging since most methods used to stain DNA (*e.g.* DAPI) are unspecific and lethal for the cells.

DNA FISH can be used to target specific loci in single cells, but can only be used in fixed cells since it requires harsh treatments and cell permeabilization to let the probes enter the nucleus and hybridize on chromatin. Multiplexed and sequential DNA FISH has been extensively used to trace chromatin fibers and to describe the heterogeneity of loop structure in single cells[89,133,134,159,160]. DNA FISH enables to obtain high localization accuracy in single cells but lacks temporal information and is subjected to fixation artifacts. However, it is a useful technique as it does not require any genome editing and can be directly performed on cells. Furthermore, given that a high number of fixed cells are analyzed, it is possible to obtain a satisfying statistical sampling of chromatin conformations.

Although fixed cell DNA labeling methods are informative, quantifying the dynamics of chromatin loop extrusion requires to record temporal information and therefore to visualize and track specific chromatin loci as function of time in living cells.

5.3.2 Methods to label DNA in living cells

Several techniques were developed to fluorescently label specific chromatin loci in living cells (**Table 1** and **Figure 14**). These techniques can be divided into two different categories: (i) inserting sequences that will be bound by fluorescent proteins, (ii) targeting fluorescent proteins to a specific sequence (**Figure 14**).

Inserting sequences bound by fluorescent proteins

Fluorescent repressor operator systems (FROS) are composed of dozens to hundreds of repetitive sequences (each spanning ~20 bp) that are known to be tightly bound by bacterial transcription factors. By fusing these transcription factors to fluorescent tags, it is possible to visualize their exogenous binding site with sufficient signal due to the dozens to hundreds of bound proteins (**Figure 14A**)[259,260]. Several orthologous systems exist (LacO, TetO, CuO[261]), enabling color multiplexing to visualize different loci simultaneously (**Table 1**)[262]. Another possibility is the ANCHOR system, which uses a 1 kb parS bacterial site to recruit ParB protein. The amplification of the signal is made by oligomerization of the ParB proteins fused to a fluorescent tag over several kbs (**Figure 14A**)[263]. Repetitive arrays are generally large (4-10 kb) compared to the ANCHOR system (< 1 kb), but both require genome editing to insert them at the desired location of the genome (**Table 1**).

Targeting fluorescent proteins to a specific sequence

The second type of live-cell chromatin methods consist in directly recruiting proteins to a specific sequence. Initially, Transcription Activator-Like Effector (TALE) proteins were designed specifically to target the sequence of interest (**Figure 14B**). Generally restricted to repetitive sequences to increase the fluorescent signal, and because TALEs had to be designed for each targeted sequence, TALEs were replaced by the use of catalytically-inactive dead Cas9 (dCas9). Cas9 proteins can be targeted to virtually any sequence in the genome by a single-guide RNA (sgRNA) (**Figure 15**). While Cas9 was initially used for genome editing due to its cutting ability, the catalytically inactive version allows to target specific regions of the genome without inducing DNA breaks[264].

Hence, by targeting the region of interest with sgRNA(s) and fluorescently labeling the dCas9 molecule, it is possible to visualize specific genomic loci in living cells[265]. Numerous studies have improved this initial method to enhance its signal by labeling the

sgRNA through MS2 and PP7 loops[266,267] or RNA-binding domains[268], using unique arrays of sgRNA sequences[269] or orthologous Cas9 for multicolor imaging[270] (**Figure 14C**). Despite these extensive improvements, dCas9 imaging is still difficult to use to image non-repetitive regions of the genome and suffers from heterogeneous fluorescent intensities (**Table 1**)[269,270].

Method	Advantages	Disadvantages	Genome editing
FROS (TetO, LacO, CuO)	- Signal can be tuned based on the number of repeats - Orthologous systems (allows multicolor imaging)	- Repetitive sequences can recombine - Long arrays might alter chromatin motion - Possible disruption of chromatin state	- Required, difficult due to long arrays
ANCHOR	- Small sequence - High signal - Orthologous systems (allows multicolor imaging)	- No control over oligomerization length (signal-to-noise ratio is difficult to tune) - Protein oligomerization might alter chromatin motion - No orthologous system - Possible disruption of chromatin state	- Required, small insertions
TALE	- No genome editing	- Needs to be designed for each sequence - Mostly restricted to repetitive sequences	- No
dCas9	- No genome editing - Orthologous systems (allows multicolor imaging) - Small sequence (minimal invasiveness)	- Off-targets - Still difficult to use on non-repetitive loci - R-loop formation	- No (except for CRISPR-tag[269])

Table 1: Comparison of live-cell chromatin labeling techniques.

These methods all have their own advantages and drawbacks (**Table 1**). Their impact of on chromatin is not known. For instance, long repetitive arrays and/or ANCHOR-based protein oligomerization might alter chromatin motion by stiffening chromatin (thereby affecting loop extrusion since it is a tension-dependent process[113]) and could also alter chromatin state (*e.g.* by inducing CpG methylation). Moreover, cells may have trouble to replicate repetitive arrays, which could also recombine and progressively lose repeats[244]. These potential experimental artifacts need to be assessed to ensure that the visualization method is minimally invasive.

Inserting sequences at precise genomic loci to visualize specific sequences (*e.g.* chromatin loop anchors) in living cells requires the use of CRISPR-Cas9, a useful tool for targeted genome editing.

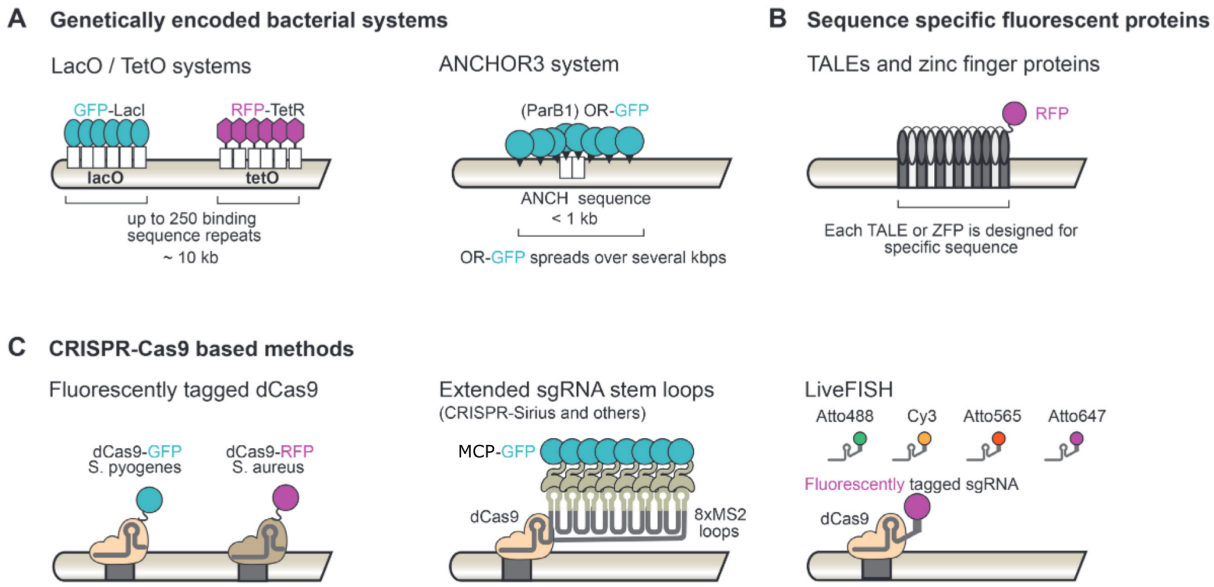


Figure 14: Techniques to label specific DNA loci in living cells. A: Repetitive short sequences (or FROS) or a *parS* sequence are inserted in the genome at the target locus. They are bound by many fluorescently labelled proteins (repetitive sequences) or the fluorescently labelled proteins oligomerize around chromatin over several kb (ANCHOR). **B:** Fluorescently labelled TALEs are specifically designed to bind to a defined sequence. **C:** dCas9 is directly recruited by its sgRNA to the targeted locus. Several improvements were made to increase the signal, such as adding MS2 or PP7 loops to the sgRNA (middle). Most dCas9-based methods do not require genome editing. Adapted from [260].

5.3.3 CRISPR-mediated genome engineering

Cas9 is a baceterial enzyme that can cut regions of the genome by being targeted to a specific genome location through a 20 bp sgRNA[271,272]. Cas9 will cut just before the PAM sequence (NGG for *Streptococcus pyogenes* Cas9), which is needed to design the sgRNA. Once a DSB has been induced at the desired location, the cellular DNA repair machinery is used to insert, delete or mutate the targeted genomic region[271].

The cell can then repair the DSB in two different ways: (i) non-homologous end joining (NHEJ) and (ii) homology-directed repair (HDR) (**Figure 15**). NHEJ can add or delete one or a few bp at the DSB site and is generally used to mutate sites. However by HDR, the DNA repair machinery searches for homology in the neighboring sequences to

recombine homologous sequences, consequently, it can be used to insert long sequences in the genome[273]. Therefore, by providing the cell with a DNA repair cassette containing the desired sequence surrounded by homology arms, it is possible to insert it specifically at the desired location. Hence, by expressing the Cas9 enzyme, the sgRNA and providing a repair template containing for instance a TetO array, it is possible to fluorescently label specific regions of the genome.

The discovery of CRISPR-Cas9 as an extremely easy and useful tool to perform genome editing opened numerous possibilities to study the genome[264]. However, a few minor drawbacks of CRISPR-Cas9 emerged. The knock-in rate is highly dependent on the location of the insertion and the size of the insert[274] and can vary from 0 to 80%[275–277]. It can be modulated by chemical modifications of the donor DNA[277] and the DNA template source (PCR, plasmid, single-stranded oligos)[278]. Finally, insertions can happen at undesired location, which is known as off-target insertions[279].

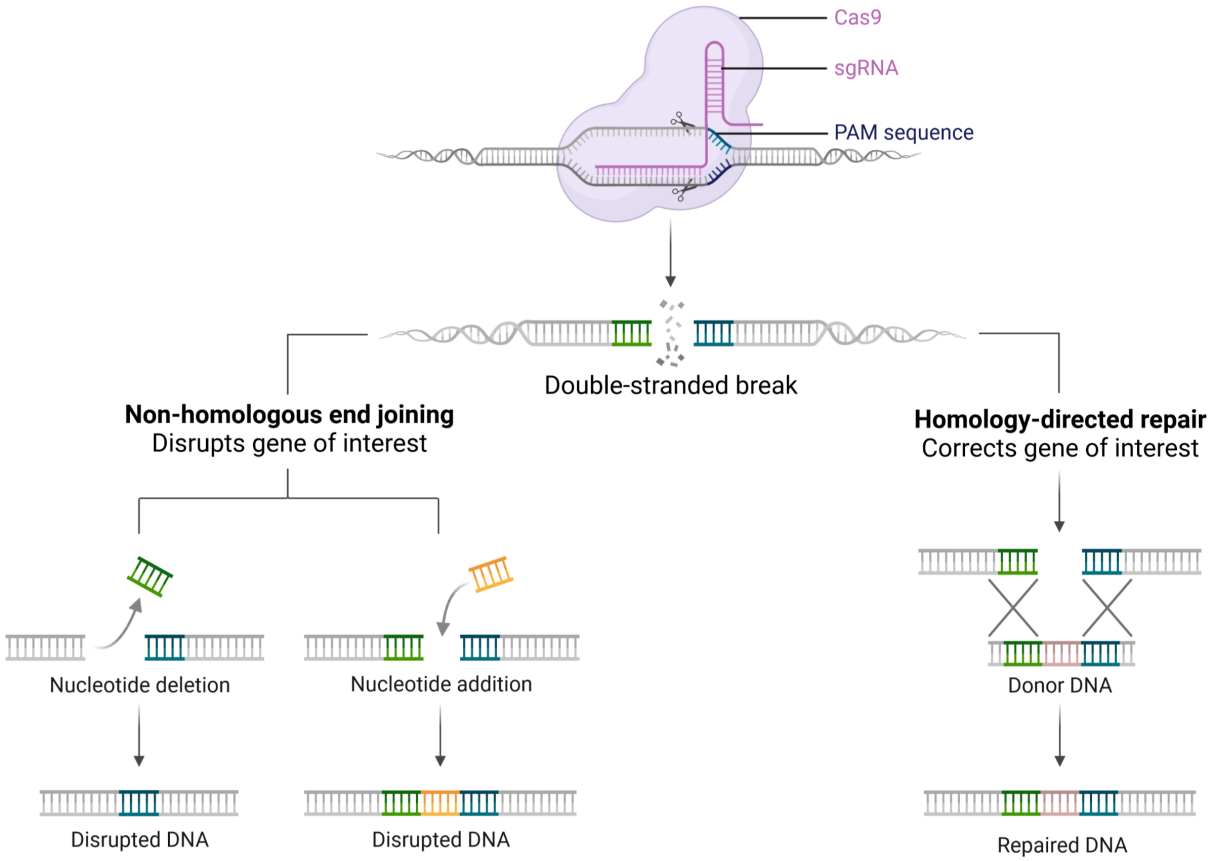


Figure 15: CRISPR genome-editing outcomes. The Cas9 enzyme is targeted to a specific locus by its sgRNA. Cas9 cuts DNA before the PAM sequence, generating a double-strand break. The cellular DNA repair machinery can repair the break using either non-homologous end joining (NHEJ, left) or homology-directed repair (HDR, right). NHEJ creates deletions,

additions of one to a few bp and can be used to induce mutations. HDR is used to insert desired long sequences at the target locus, thanks to the homology arms (green and blue sequences) surrounding the sequence to insert (grey). Figure made with Biorender.com.

5.4 First evidence of the dynamic nature of chromatin loops in living cells

During my PhD, two groups published results of chromatin loop anchors tracking in living mESCs. A full discussion of their results is made in section **2. Chromatin loops are consistently dynamic across model systems** and **Figure 20**.

The two studies compared the distance between chromatin loop anchors in conditions where cohesin or CTCF were mutated, as well as conditions where CTCF sites were deleted. From these experimental data, together with modeling approaches, they were able to demonstrate that chromatin loops were rarely found in a 'closed' state where both anchors are in close proximity. Quantitatively, loops were found in the closed state 3-6%[164] or 27%[163] of the time with mean lifetimes of 5-15 min[163] and 15-45 minutes[164]. Moreover, TADs were found to be undergoing extrusion 92% of the time[164], arguing in favor of a model where cohesin constantly extrudes chromosomes but only rarely stalls at loop boundaries. Also, by comparison with polymer simulations, they were able to estimate loop processivity (150 kb) and loop density (1 cohesin/240 kb) with comparable results to previous estimates[32], as well as the probability of CTCF to stall cohesin (12.5%)[164].

The results of the two studies are interestingly consistent with each other, considering the differences in experimental designs. One of the two studies targeted a smaller (150 kb), synthetic loop created by the insertion of a triplet of strong CTCF sites[163], while the other targeted a large (505 kb) endogenous TAD[164]. Accordingly, the fractions and fractions of closed states were higher in the smaller loop than in the bigger loop. However, none of these studies quantified the *in vivo* loop extrusion rate, as well as the heterogeneity between different loops and TAD structures. Moreover, these studies quantified chromatin loop dynamics in mESCs, which are known to display a more dynamic and accessible genome than differentiated cells[280–282].

The two above-mentioned studies have shed first light on the dynamics of chromatin loop anchors in mESCs. Each of these studies focused on a single genomic locus with a defined genomic length and chromatin environment. However, the size of chromatin loops

can go from ~100 kb up to 1 Mb and more than 10,000 loops are identified genome-wide, each harboring their unique chromatin environments[26,27]. As discussed in this introduction, this variability in chromatin environments is likely to affect chromatin loop dynamics and was not captured by these studies. Also, a significant part of the estimations made in these papers was based on the comparison between polymer simulations and experimental data. Although such approach allowed the authors to estimate otherwise masked parameters, models cannot capture the whole biological complexity. Hance, results emanating from the comparison between polymer simulations and experimental data should be interpreted with caution. A fuller and more empirical characterization of loop extrusion, across multiple genomic loci is still lacking.

6. Objectives of the PhD

6.1 Multiple-loci quantification of loop anchor dynamics in living human cells

This introduction showed that cohesin-mediated loop extrusion in the G1 phase of the cell cycle is tightly regulated. Protein partners, RNA, chromatin states and DNA sequence (through the location, distribution and affinity of CTCF binding sites) all contribute to modify cohesin dynamics on chromatin. Loop extrusion dynamics is involved in the regulation of multiple key nuclear processes (*e.g.* gene expression regulation or DNA repair) that guarantee genome integrity and is conserved across evolution. Despite the seemingly great importance of this mechanism for the cell physiology, the dynamics of chromatin loops, the structures emanating from loop extrusion, is yet partially characterized.

During my PhD, I aimed at quantifying chromatin loop dynamics in living human cells. Using CRISPR genome editing and live-cell fluorescent labeling of chromatin, as well as conditional depletion of cohesin, we could efficiently visualize and quantify the molecular process of loop extrusion. Using dedicated analysis methods tested on polymer simulations, we estimated the effective loop extrusion speed directly from experimental data. We then compared chromatin loop dynamics of multiple genomic loci to characterize the variability of loop anchor dynamics. This allowed us to determine whether common parameters of loop extrusion can explain the dynamics of chromatin loops at different loci.

6.2 Motivating questions

The goal of my PhD was to visualize and quantify chromatin loop extrusion in living human cells. To achieve this goal, I wanted to answer the following questions:

- Can loop extrusion be detected and quantified by fluorescently imaging loop anchors, despite the unavoidable corruption of experimental data by various sources of noise?
- Which methods can I use to estimate biophysical parameters of loop extrusion from fixed and live-cell imaging of loop anchors?
- How long and frequent are the interactions between loop anchors?
- What is the effective speed of loop extrusion in living cells?
- What is the variability of chromatin loop dynamics between different genomic loci and can genomic features explain this variability?

6.3 Computational and experimental quantification of chromatin loop extrusion dynamics

To address these questions, I made use of a wide panel of techniques ranging from *in silico* polymer modeling to live-cell imaging. I detail below the different necessary steps to achieve the quantification of chromatin loop extrusion dynamics in living human cells.

- The use of fluorescent labels to track chromatin loop anchors has a major drawback: one can only follow the loop anchors, while the rest of the loop structure is hidden. Due to this lack of information, together with the stochastic movements of the chromatin polymer and the unavoidable noise in experimental data, it is not *a priori* obvious whether it is possible to detect and quantify loop extrusion by imaging loop anchors. Therefore, I turned to polymer simulations where the known ground truth allowed me to test the experimental conditions in which loop extrusion can be detected and to validate analysis methods to quantify this process.
- The visualization of chromatin in living cells had to be optimized while keeping the invasiveness of the method minimal. I developed molecular cloning strategies to efficiently create numerous CRISPR repair cassettes and optimized CRISPR-mediated insertion rates. Finally, I optimized live-cell imaging conditions to enable long and accurate tracking of loop anchors.

- I could visualize the loop extrusion process at three different regions of the genome, each displaying unique loop strengths, genomic distances and epigenetic profiles. By using the methods developed and validated on the polymer simulations, I could quantify and compare loop extrusion dynamics across different genomic loci, with or without loop extrusion.

In this manuscript, I present the different modeling and experimental results in the following order:

- Using polymer simulations, I defined guidelines about how to detect loop extrusion experimentally and estimate meaningful biophysical parameters from chromatin loop imaging (Results section **1. Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging**).
- The experimental design and optimization of the live-cell imaging method (Results section **2. Tracking loop anchors in living cells**).
- The visualization and quantification of chromatin loop extrusion by imaging multiple loops and TADs in living cells (Results section **3. Visualization and quantification of chromatin loop extrusion in living human cells**).

Results

1. Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging

1.1 Can we detect and quantify chromatin loop extrusion by fluorescently imaging loop anchors?

To quantify chromatin loop extrusion dynamics in living cells, I decided to label loop anchors fluorescently, each with a different color. However, the fluorescent labeling of loop anchors requires sequential rounds of CRISPR-Cas9 genome editing and cell cloning. It is a tedious and long process to build and obtain such cell lines. Before rushing into time-consuming experiments, I used polymer simulations to assess the experimental feasibility and define the experimental conditions under which loop extrusion can be detected and quantified by tracking loop anchors fluorescently (**Figure 1** of Sabaté *et al*, NAR, 2023)[101].

Molecular dynamics simulations are used to model molecular processes and interactions *in silico*. They constitute a powerful tool, in particular when combined with experiments. The computational modeling of molecular processes allows to test an extensive number of parameters and a wide range of conditions at low temporal and economical costs. By modeling molecular processes (*e.g.* chromatin loop extrusion), one can gain access to data that are experimentally limited or difficult to obtain, test hypothetical mechanisms, validate analysis methods prior to experimental design, define methods' limitations and the impact of various sources of noise on the inferred results.

From the experimental tracking of fluorescently labelled loop anchors, I aim to quantify the loop extrusion process and its dynamics. In principle, this goal appears straightforward. As the loop is enlarged by the moving cohesin complex, the distance between the two loop anchors is expected to decrease. The anchor-anchor distance then becomes minimal when the cohesin complex reaches the two anchors and stalls. However, numerous complications arise from the biological and experimental sources of uncertainty (*e.g.* finite localization precision (standard deviation of localization errors) in noisy images and stochastic movements of the chromatin fiber). Furthermore, the anchor-anchor distance constitutes a limited amount of information since I lack information about the whole loop

structure. Hence, it is *a priori* unknown to what extent loop extrusion can be detected and quantified by labeling loop anchors. I have chosen to use polymer simulations and model the loop extrusion process *in silico*, in order to evaluate the experimental and technical conditions enabling the detection of loop extrusion by measuring the loop anchor distance (**Figure 1** of Sabaté *et al*, NAR, 2023).

Live-cell tracking of two genomic loci separated by a few hundred kbs has not been widely used to study genome architecture[163,164,283–285]. Because of this novelty, few analytical methods have been developed to analyze such data. Using polymer simulations where the ground truth is known, I tested and validated analytical methods to quantify loop extrusion dynamics from static and dynamic imaging of loop anchors (**Figure 1** of Sabaté *et al*, NAR, 2023). Namely, I developed methods to:

- Use static imaging alone to estimate the fraction of loops in three different states: open (completely unextruded), extruding (actively extruding), and closed (fully extruded) state.
- Segment the closed states, quantify their lifetime and the effective speed of loop extrusion from live-cell dynamic imaging.

Thus, the use of polymer simulations guided the experimental design and data analysis, allowing to maximize the amount of information retrieved from such experiments. Simulations allowed us to prove that loop extrusion can be detected by imaging loop anchors alone, despite numerous sources of noise. Finally, I could define what information can be estimated from static or dynamic imaging, as well as test and validate analytical methods to analyze this novel type of data.

2614–2632 *Nucleic Acids Research, 2023, Vol. 51, No. 6* *Published online 25 February 2023*
https://doi.org/10.1093/nar/gkad034

Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging

Thomas Sabaté [1,2,3,*], Benoît Lelandais[1,4], Edouard Bertrand[2,†] and Christophe Zimmer[1,*,†]

[1]Institut Pasteur, Université Paris Cité, CNRS UMR 3691, Imaging and Modeling Unit, F-75015 Paris, France, [2]IGH, University of Montpellier, CNRS UMR 9002, Montpellier, France, [3]Sorbonne Université, Collège Doctoral, F-75005 Paris, France and [4]Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France

Received June 02, 2022; Revised December 03, 2022; Editorial Decision January 03, 2023; Accepted January 25, 2023

ABSTRACT

Genome-wide chromosome conformation capture (Hi-C) has revealed the organization of chromatin into topologically associating domains (TADs) and loops, which are thought to help regulate genome functions. TADs and loops are understood as the result of DNA extrusion mediated by the cohesin complex. However, despite recent efforts, direct visualization and quantification of this process in single cells remains an open challenge. Here, we use polymer simulations and dedicated analysis methods to explore if, and under which conditions, DNA loop extrusion can be detected and quantitatively characterized by imaging pairs of fluorescently labeled loci located near loop or TAD anchors in fixed or living cells. We find that under realistic conditions, extrusion can be detected and the frequency of loop formation can be quantified from fixed cell images alone, while the lifetime of loops and the speed of extrusion can be estimated from dynamic live-cell data. Our delineation of appropriate imaging conditions and the proposed analytical methods lay the groundwork for a systematic quantitative characterization of loop extrusion in fixed or living cells.

INTRODUCTION

Over the last decade, much progress has been made in understanding the three-dimensional organization of chromatin, thanks to powerful genomic techniques such as Hi-C (1), which provides genome-wide maps of DNA–DNA contact frequencies (2,3). A notable milestone was the discovery of topologically associating domains (TADs) (4,5), sub-megabase scale regions of enhanced chromatin contacts that appear as blocks on the diagonal of Hi-C maps and that are believed to help regulate gene expression by facilitating - or on the contrary impeding - enhancer-promoter interactions (6–15). TADs are often associated with off-diagonal peaks (or 'corner dots') in the Hi-C maps, reflecting enriched contacts between the two distant loci that define the TAD boundaries (hereafter called 'anchors'), and which are interpreted as chromatin loops. TAD or loop anchors are typically binding sites for the insulator protein CTCF (1) (with convergent orientation of the CTCF motifs) and TADs with corner dots (also called loop domains) depend on the ring-like cohesin complex, as they disappear upon cohesin removal (16). The formation of TADs and corner dots is now understood as the result of DNA loop extrusion mediated by the cohesin complex (2,3,17–20). In this process, the cohesin ring complex binds to DNA and progressively pulls out a loop of chromatin, until the complex unbinds or stops extruding at obstacles such as CTCF-bound anchor loci. At this point, the loop is temporarily stabilized until the cohesin complex or CTCF dissociate and the anchors detach from each other (21,22). This extrusion mechanism is supported by several lines of evidence, including polymer modeling (19,20,23), Hi-C studies where cohesin (16), CTCF or other regulators of the cohesin complex are experimentally depleted (18,24,25), as well as direct visualization of cohesin-mediated DNA extrusion *in vitro* (17,26). The dynamic nature of TADs and loops is further supported by multiplexed DNA FISH studies, which underlined the high cell-to-cell heterogeneity in chromatin structure within TADs or loops (27–32), and by single-molecule tracking of CTCF and the cohesin subunit RAD21, which showed that these factors have residence times on chromatin orders of magnitude shorter than the cell cycle (1–2 min for CTCF and 22 min for RAD21) (33).

Despite these studies, visualizing and characterizing the dynamic process of loop extrusion directly in single living cells remains a largely unaddressed challenge (34), except for two very recent reports (35,36). A seemingly straightforward experimental approach to visualize loop extrusion in living cells is to track two loop anchors with light

*To whom correspondence should be addressed. Tel: +33 140613170; Email: thomas.sabate@pasteur.fr
Correspondence may also be addressed to Christophe Zimmer. Tel: +33 140613891; Email: czimmer@pasteur.fr
†The authors wish it to be known that, in their opinion, the last two authors should be regarded as Joint Last Authors.

microscopy, using distinct fluorescent reporters located at or near each anchor, and to monitor the progressive decrease of the reporter-reporter distance that is expected to result from extrusion (37). Fluorescent labeling of anchors can be achieved using DNA FISH probes in fixed cells (28), while arrays of repeats (38), dead Cas9 (39–41) or parS-parB (42) can be used to label loci both in fixed and living cells. In practice, however, direct visualization of extrusion by tracking fluorescent loci is complicated by several biological and experimental sources of uncertainty. These include: (i) unavoidable errors in computing spatial coordinates and distances between genomic loci from noisy imaging data, (ii) photobleaching, which limits the number of time points over which loci can be tracked with sufficient signal intensity for accurate localization, (iii) the size of fluorescent labels and the distance of fluorescent reporters to the anchors (hereafter called reporter–anchor separation), (iv) the stochastic movements of chromatin, which can bring together genomically distant loci in space even in absence of any active process such as loop extrusion, (v) the potential rarity of extrusion events and by consequence the potentially large number of cells that must be analyzed to accurately characterize statistical parameters such as the average lifetime of loops. Because of these complicating factors, it is not *a priori* evident whether loop extrusion can be unambiguously visualized by imaging at all, and if it can, under what experimental conditions, and whether key biophysical parameters of loop extrusion can be quantified. Here, we aim to clarify these requirements considering basic expectations from polymer dynamics, taking into account available Hi-C data, and considering various technical limitations of imaging techniques. We approach this by (i) simulating realistic distributions and time series of anchor–anchor distances, (ii) proposing analytical methods to characterize loop extrusion from these data and (iii) quantitatively testing these methods on the simulations. Our results will guide future experimental work aiming to quantify loop extrusion and its dynamics in fixed and living cells.

MATERIALS AND METHODS

Our study involves (i) numerical simulations of polymers undergoing loop extrusion, (ii) analytical models of probability distributions of anchor–anchor vectors and distances, (iii) analyses of simulated data, and (iv) analysis of experimental data from Hi-C, ChIP-seq or imaging. The following provides details on these four methodological parts.

Polymer simulations with loop extrusion

Langevin dynamics simulations. We used polymer simulations to model the dynamics of a chromatin fiber segment subjected or not to loop extrusion. The simulated polymer consisted of 600 beads, and polymer motions were simulated with Langevin dynamics in fixed boundary conditions using LAMMPS (43). Consecutive beads were connected by a harmonic bond with a potential $E_{\text{bond}} = 30(r-1)^2$, where r is the distance between bead centers. The polymer stiffness was modeled using a harmonic potential $E_{\text{bending}} = K_0\,(\theta - \theta_0)^2$, where K_0 is the stiffness parameter (set to $K_0 = 0.1$), θ is the angle between three consecutive beads and $\theta_0 = 180^\circ$ is the equilibrium value (corresponding to three aligned beads). We verified that the contact frequencies did not strongly depend on the exact value of K_0 (see section '*Simulated contact frequency maps*' below). By default, the polymer was confined in a sphere of radius 18 bead diameters using the energy potential $E_{\text{wall}} = 4[(\frac{\sigma}{R})^{12} - (\frac{\sigma}{R})^{6}]$ for $R < R_c$, where R is the distance between the surface of the confining sphere and the center of a bead, σ is a size factor set to 0.5 bead diameters and R_c is the cutoff distance set to 0.5 bead diameters (for $R \geq R_c$, $E_{\text{wall}} = 0$). This confinement implied a volume occupancy ratio of the polymer of $\rho = 1.3\%$. While this is much lower than estimates of chromatin volume occupancy *in vivo* (∼10–15%) (44–46), this discrepancy is not critical given the consistency of our simulation predictions with Hi-C and imaging data (see Results sections 'Comparing polymer simulations to Hi-C data' and 'Comparing polymer simulations to imaging data'). We considered the polymer to be equilibrated when its radius of gyration and end-to-end distance were both stabilized, which was the case after ∼7.5 million time steps (Supplementary Figure S1A, B). After this, we recorded the positions of each bead every 1000 time steps until the end of the simulation (≈12 million time steps). Polymer coordinates were imported for further analysis in Python using MDAnalysis (47).

In order to convert simulation time and space dimensions to physical units, we compared Mean Squared Displacement (MSD) from simulations to experimental MSD curves of chromatin loci tracked by live-cell microscopy (40,41,48). This comparison led to the conversion of 1000 simulation time steps to 0.3 s and of 1 bead diameter to $2r_0 = 50$ nm. Assuming a chromatin compaction of $C = 60$ bp/nm (49), as previously estimated by comparing simulation predictions to experimental data in yeast, this implies that 1 bead corresponds to $g = 2r_0C = 3$ kb and the entire 600 bead polymer to 1800 kb of chromatin.

Simulating loop extrusion. To model loop extrusion, we forced the formation of a harmonic bond between non-consecutive beads. We assumed that loop extrusion initially occurs bidirectionally (17,26), i.e. that if beads i and j ($i < j$) are bonded at time t, then beads $i-1$ and $j+1$ are bonded at time $t + \Delta t$ (while the bond between beads i and j is deleted), where Δt is the time needed to extrude two beads (Supplementary Figure S2A). The speed of loop extrusion (in base pairs per seconds) is thus defined as $V_0 = 2g/\Delta t$. By default, our simulations assumed that extrusion started at a random location between the beads representing the anchors (extrusion barriers) and proceeded bidirectionally at $V_0 = 1$ kb/s (17,26) until reaching an anchor. Thereafter, loop extrusion proceeded unidirectionally, at the halved speed $V_0/2 = 0.5$ kb/s until reaching the second anchor (18,50–52), whereupon extrusion stopped (Supplementary Figure S2A). By default, we then maintained the bond between the two anchor beads until the end of the simulation. However, for simulations used in the Results sections 'Quantifying closed loop lifetimes from live-cell trajectories' and 'Quantifying the speed of loop extrusion from live-cell trajectories', the time spent in the closed state (i.e. when the anchors are maintained in contact by a bond) was drawn from a truncated exponential distribu-

tion. At the end of the closed state, the bond linking the two anchors was deleted and the polymer was allowed to relax without loops until the end of the simulation. We simulated loops ranging from 150 kb to 990 kb in size, and defined the positions of the anchor beads using the following bead indexes: 150 kb loops: {275, 324}, 228 kb loops: {262, 337}, 300 kb loops: {250, 349}, 450 kb loops: {225, 374}, 504 kb loops: {216, 383}, 600 kb loops: {200, 399}, 702 kb loops: {183, 416}, 798 kb loops: {167, 432}, 990 kb loops: {135, 464}.

For Supplementary Figure S11, and unlike elsewhere in the paper, we performed simulations where extrusion always started in the middle of the loop and proceeded bidirectionally at constant speed, until both anchors were reached simultaneously. This was done to ensure that extrusion started and ended at the same time in all cases, which facilitated the comparison of simulations with the theoretical linear model (see Eq. (3) and Supplementary Figure S10B, green dashed line). For each value of these parameters, four hundred synchronized time series were averaged together. We simulated loops of 150 kb, 300 kb, and 600 kb extruding at speeds of 0.2, 1 and 5 kb/s, and loops of 990 kb extruding at speeds of 1 and 5 kb/s. We also repeated these simulations with much weaker confinement (using a sphere of radius 150 rather than 18 bead diameters).

Analytical models

Anchor–anchor vector distributions. In the Results section 'Estimating the fraction of loop states from static imaging data', we used an analytical model to estimate the fractions of loops in open vs extruding vs closed states (these states correspond respectively to absence of loops, loops whose size increases with time, and to a stable loop with the two anchors in contact) based on the measured coordinates of anchor–anchor vectors. This model is based on the basic properties of an ideal polymer chain, for which the anchor–anchor vector $\vec{R} = (\delta x; \delta y; \delta z)$ is a random variable that obeys the normal probabilitydensity: $P\left(\vec{R};\sigma\right) = P_x(\delta x;\sigma)\, P_y(\delta y;\sigma)\, P_z(\delta z;\sigma)$ with: $P_w(\delta w;\sigma) = \left(\sqrt{2\pi}\sigma\right)^{-1} exp(-\frac{\delta w^2}{2\sigma^2})$ for each coordinate $w \in \{x, y, z\}$ and where the variance σ^2 is proportional to the number N_b of Kuhn lengths b separating the anchors and is given by $\sigma^2 = N_b b^2/3$. This implies that : $P(\vec{R};\sigma) = (2\pi\sigma^2)^{-\frac{3}{2}}\, exp(-\frac{R^2}{2\sigma^2})$ and the mean squared anchor–anchor distance (MSAAD) is given by: $\langle R^2 \rangle = N_b b^2$, where the brackets denote statistical averaging. In presence of a loop, we assume that the parts of the polymer outside the loop behave as if the part within the loop was absent, thereby shortening the number of Kuhn lengths between the anchors. Note that in the absence of bending stiffness ($K_0 = 0$), the Kuhn length coincides with the bead diameter ($b = 2r_0 = 50$ nm), which for an anchor–anchor separation of 150 kb, corresponding to $150/g = 50$ beads and 49 Kuhn lengths would imply $\sigma_{\text{open}} = \sqrt{\frac{N_b b^2}{3}} \approx 202$ nm. However, because of the bending potential E_{bending} assumed above, the variance measured on simulated data is slightly larger, namely $\sigma_{\text{open}} = 216$ nm, implying a slightly larger Kuhn length of $b = 53.5$ nm. Also note that the finite bead radius of $r_0 = 25$ nm used in our simulations affects anchor-to-anchor vectors in the closed state and at the end of the extruding phase, since distances between anchors fluctuate around $2r_0 = 50$ nm. To avoid this bias, which is not accounted for by the above analytical model, we shortened the anchor–anchor distances predicted by the simulations in the closed state by 50 nm, and between 0 and up to 50 nm for the last 60 time points of the extruding state. After this correction, in the closed state, measurement of the MSAAD led to a small value of $\sigma_{\text{closed}} \approx 7$ nm.

The number of Kuhn lengths N_b between the two anchors depends on the loop state. It is largest ($N_b = N_0$) when the loop is open, equals zero when the loop is closed, and assumes intermediate values during extrusion.

The above model for anchor–anchor vectors applies to the open state with $\sigma^2_{\text{open}} = N_0 b^2/3$ and to the closed state with $\sigma^2_{\text{closed}} \ll \sigma^2_{\text{open}}$. Hence:

$$\begin{cases} P_w^{\text{open}}\ (\delta w) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma_{\text{open}}}\ exp\left(-\frac{\delta w^2}{2\sigma^2_{\text{open}}}\right) \\ P_w^{\text{closed}}\ (\delta w) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma_{\text{closed}}}\ exp\left(-\frac{\delta w^2}{2\sigma^2_{\text{closed}}}\right) \end{cases}$$

In the extruding state, the number of Kuhn lengths continuously varies from N_0 to 0. As a result, the probability density $P_w^{\text{extruding}}(\delta w)$ is an integral over σ^2 varying from σ^2_{closed} up to σ^2_{open}. If the speed of extrusion was constant, then loops of all sizes (from $N_b = 0$ to $N_b = N_0$) would be represented equally in the integral, such that:

$$P_w^{\text{extruding}}(\delta w) = \int_{\sigma^2_{\text{closed}}}^{\sigma^2_{\text{open}}} P_w(\delta w; s) ds^2$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{\sigma^2_{\text{closed}}}^{\sigma^2_{\text{open}}} \frac{1}{s} exp(-\frac{\delta w^2}{2s^2}) ds^2.$$

A complication arises from the fact that the speed of extrusion is halved once the extrusion complex reaches one of the two anchors and extrusion switches from bidirectional to unidirectional. As a consequence, extruding states with unidirectional extrusion are twice more frequent than states with bidirectional extrusion, leading to twice more frequent cases where $N_b = 0$ than where $N_b = N_0$. To take this into account, we introduce a weighting factor $D(\sigma)$ in the integral, such that:

$$P_w^{\text{extruding}}\ (\delta w) = \int_{\sigma^2_{\text{closed}}}^{\sigma^2_{\text{open}}} D(s)\, P_w(\delta w; s) ds^2$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{\sigma^2_{\text{closed}}}^{\sigma^2_{\text{open}}} \frac{D(s)}{s} exp(-\frac{\delta w^2}{2s^2}) ds^2.$$

Assuming that extrusion is initiated at uniformly random locations between the two anchors, $D(\sigma)$ increases linearly

with σ^2 between σ^2_{open} and σ^2_{closed}, such that:

$$D(\sigma) = \frac{2\sigma^2_{\text{open}} - \sigma^2_{\text{closed}} - \sigma^2}{\frac{3}{2}\sigma^4_{\text{open}} + \frac{3}{2}\sigma^4_{\text{closed}} - 3\sigma^2_{\text{closed}}\sigma^2_{\text{open}}}$$

This equation can be derived by considering a linear function that equals 2 when $\sigma = \sigma_{\text{close}}$ and 1 when $\sigma = \sigma_{\text{open}}$ and whose integral between σ^2_{close} and σ^2_{open} equals 1 (to ensure that it is a probability density).

In practice, the coordinates of anchors (or nearby fluorescent reporters) are not perfectly known, but are computed from noisy images. Because of noise, these coordinates are estimated with a finite precision, which is usually different along the axial and lateral directions of the microscope. To reflect this, we assumed that coordinates are perturbed by random, normally distributed anisotropic errors (hereafter called localization errors), with standard deviations σ_w along each axis $w \in \{x, y, z\}$. This leads to the following modified equations for the probability densities of anchor–anchor coordinate differences in each of the three loop states:

$$\begin{cases} P_w^{\text{open}}(\delta w) = \frac{1}{(2\pi)^{\frac{1}{2}}\sqrt{\sigma^2_{\text{open}}+2\sigma^2_w}}\, exp\left(-\frac{\delta w^2}{2\sigma^2_{\text{open}}+4\sigma^2_w}\right) \\ P_w^{\text{closed}}(\delta w) = \frac{1}{(2\pi)^{\frac{1}{2}}\sqrt{\sigma^2_{\text{closed}}+2\sigma^2_w}}\, exp\left(-\frac{\delta w^2}{2\sigma^2_{\text{closed}}+4\sigma^2_w}\right) \\ P_w^{\text{extruding}}(\delta w) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{\sigma^2_{\text{closed}}}^{\sigma^2_{\text{open}}} \frac{D(s)}{\sqrt{s^2+2\sigma^2_w}} \exp\left(-\frac{\delta w^2}{2s^2+4\sigma^2_w}\right) ds^2 \end{cases} \quad (1)$$

For a combination of the three states with fractions A_{closed}, A_{open} and $A_{\text{extruding}} = 1 - A_{\text{closed}} - A_{\text{open}}$, the full probability density of anchor–anchor coordinate differences reads:

$$P_w(\delta w) = A_{\text{closed}} P_w^{\text{closed}}(\delta w) + A_{\text{open}} P_w^{\text{open}}(\delta w) + A_{\text{extruding}} P_w^{\text{extruding}}(\delta w) \quad (2)$$

Mean anchor–anchor distance as function of time. In the Results section 'Quantifying the speed of loop extrusion from live-cell trajectories', we analyzed time series of anchor–anchor distances using an analytical model of the MSAAD $\langle R^2\rangle(t)$ as function of time. For a given time series, we defined $t = 0$ as the time point when extrusion stops, i.e. the start of the closed state. We further assumed, as above, that the MSAAD obeys the properties of an ideal polymer chain whose length (i.e. its number of monomers) is diminished by the length of the loop. Accordingly, if extrusion proceeded at a constant speed V_0, then the number of Kuhn lengths between the two anchors would decay linearly with time as $N_b(t) = N_0(V_0/s_0)(-t)$ for $t \in [-\frac{s_0}{V_0}, 0]$, where N_0 is the number of Kuhn lengths in the open state and s_0 is the genomic distance between the anchors, in base pairs.

Under the assumptions above, and ignoring localization errors, we have $\langle R^2\rangle(t) = N_b(t)b^2 = N_b(t)/N_0 R_0^2$, where R_0^2 is the MSAAD in the open state (before extrusion starts). Because of random localization errors, the measured MSAAD is $\langle R^2\rangle(t) = N_b(t)/N_0 R_0^2 + R^2_{\text{loc}}$, where $R^2_{\text{loc}} = 4\sigma^2_{x,y} + 2\sigma^2_z$ is the contribution of random localization errors.

Thus, for a constant extrusion speed V_0, the MSAAD obeys the following linear law:

$$\langle R^2\rangle(t) = R_0^2(V_0/s_0)(-t) + R^2_{\text{loc}} \quad (3)$$

However, we assumed that loop extrusion switches from bidirectional to unidirectional, and the speed of extrusion changes from V_0 to $\frac{V_0}{2}$, which leads to a non-linear dependence of the MSAAD with time (Supplementary Figure S10A). For any given set of time series with the same extrusion initiation site, the MSAAD is bounded by R_0^2 and $d_1^2 = R_0^2\,\frac{V_0(-t)}{2s_0}$ for times $t \le -\frac{s_0}{V_0}$, whereas it is bounded by d_1^2 and $d_2^2 = R_0^2\,\frac{V_0(-t)}{s_0}$ for times $t \ge -\frac{s_0}{V_0}$ (Supplementary Figure S10B).

We now denote as $\langle\langle R^2\rangle\rangle(t)$ the ensemble average of the MSAAD (EMSAAD) over many time series, still assuming that $t = 0$ is the start of the closed state for all time series. The theoretical EMSAAD can be derived under the assumption that extrusion is initiated with uniform random probability between the two anchors. With this assumption, for times $t \le -\frac{s_0}{V_0}$, the MSAAD equals R_0^2 with probability $p_1(t) = -\frac{3V_0t}{4s_0} - \frac{1}{2}$ and equals d_1^2 with probability $1 - p_1(t)$ (Supplementary Figure S10C). Thus, for $t \le -\frac{s_0}{V_0}$, we have: $\langle\langle R^2\rangle\rangle(t) = p_1(t)R_0^2 + (1 - p_1(t))d_1^2$. For times $t \ge -\frac{s_0}{V_0}$, the MSAAD equals d_2^2 with probability $p_2(t) = -\frac{V_0t}{4s_0}$ and equals d_1^2 with probability $1 - p_2(t)$ (Supplementary Figure S10C), hence $\langle\langle R^2\rangle\rangle(t) = p_2(t)\,d_2^2 + (1 - p_2(t))d_1^2$. This leads to a complete, parameter-free model for the EMSAAD at all times:

$$\langle\langle R^2\rangle\rangle(t) = \begin{cases} -R_0^2\left(\frac{3V_0t}{2s_0}\left(\frac{V_0t}{4s_0}+1\right)+\frac{1}{2}\right) + R^2_{\text{loc}} & \text{if } t < t_0 \\ R_0^2\,\frac{V_0t}{2s_0}\left(\frac{V_0t}{4s_0}-1\right) + R^2_{\text{loc}} & \text{if } t_0 \le t \le 0 \end{cases} \quad (4)$$

where $t_0 = -s_0/V_0$. Note that for $t = t_0$, the EMSAAD is: $\langle\langle R^2\rangle\rangle(t = t_0) = \frac{5}{8}R_0^2 + R^2_{\text{loc}}$.

Analysis of simulated data

Simulated contact frequency maps. To predict chromatin contact frequency maps for comparison with Hi-C data, we used ensembles of simulated polymer conformations as follows. For simulations of 300 kb loops or larger, we generated 2500 independent simulations and randomly picked 80 polymer conformations per trajectory, yielding 200 000 single conformations in total. For 150 and 228 kb loops, we randomly picked 50 conformations from 4000 independent simulations, also resulting in 200 000 single conformations in total. Predicting a contact map from an ensemble of polymer conformations requires to define a capture radius (i.e. a threshold for the spatial distance between bead centers below which any pair of beads generates a contact event). To determine this radius, we first computed contact maps at 3 kb resolution for various radii, then calculated the average contact frequency along each diagonal of the contact map as function of genomic separation, $P(s)$. We compared simulated $P(s)$ curves to the experimental Hi-C counterpart over the range $s \in [5, 300]$ kb. We achieved the best match for a capture radius of 3 beads (150 nm) (Supplementary Figure S2B). This radius agrees with distance thresholds used in previous studies to reproduce Hi-C maps

from distances between pairs of loci measured by DNA FISH (30,31). As mentioned above (section '*Langevin dynamicssimulations*' above), we also verified that changes in the stiffness parameter K_0 around its assumed default value of 0.1 did not strongly affect simulated $P(s)$ (Supplementary Figure S2D). Contact frequency maps were normalized by the total number of conformations.

Contact peak scores. In order to quantitatively compare the strength of loops in simulated contact maps and Hi-C data, we defined a peak score as the contact frequency between the two anchors divided by a background contact frequency (Figure 2B). For this purpose, the simulated contact maps and the selected Hi-C contact maps (see section '*Analysis of ChIP-Seq and Hi-C data*' below), with genomic resolutions of 3 and 5 kb, respectively, were re-binned to a common resolution of 15 kb and were normalized to a sum of 1. The background frequency was defined as the average contact frequency inside a 30 × 30 kb (i.e. 2 × 2 bin) window located at 30 kb from each anchor and inside the loop domain (Figure 2B). Peak scores of simulated or experimental (Hi-C) contact frequency maps were compared using a one-sample t-test.

Quantification of loop extrusion detectability. The ability to detect loop extrusion from anchor–anchor distances measured in fluorescence microscopy images depends on different parameters that include loop size, extrusion speed, localization precision, reporter–anchor separation, fractions of loop states (open, extruding and closed) and the sample size. To assess the ability to detect loop extrusion for a given set of parameters, we randomly picked N anchor–anchor distances from 1000 independent simulation trajectories, each containing 3400 time points, with or without extrusion (Figure 3A). We compared cumulative distance distributions with and without extrusion ($F(x)$ and $G(x)$, respectively) using one-sided Kolmogorov–Smirnov tests (with the alternative hypothesis $F(x)<G(x)$). This test was repeated on 5000 independent random samples of the two distance distributions. We defined detectability of loop extrusion as the percentage of Kolmogorov–Smirnov tests revealing significant differences ($P < 0.05$ after Benjamini–Hochberg correction) (53) (Figure 3B).

To mimic reporter–anchor separation (Figure 3C), we measured the distances between beads that were shifted relative to the beads representing loop anchors. To mimic localization errors (Figure 3D), we shifted the (x, y, z) coordinates of the tracked bead centers using random normally distributed displacements $(\delta x, \delta y, \delta z)$ of mean 0 and standard deviations $\sigma_x = \sigma_y = \sigma_{x,y}$ and $\sigma_z = 2\sigma_{x,y}$, respectively, where the factor 2 reflects the typical anisotropy due to the axial elongation of widefield or confocal point spread functions (54). To model different loop extrusion speeds (Figure 3H), we modified the time interval Δt after which new bonds were created during loop extrusion (see section '*Simulating loop extrusion*' above).

Fitting theoretical distance models. In the Results section 'Estimating the fraction of loop states from static imaging data', we used the theoretical model from (Eq. 2) above to estimate the three state fractions A_{closed}, A_{open} and $A_{\text{extruding}} = 1 - A_{\text{closed}} - A_{\text{open}}$ from simulated anchor–anchor vectors. We randomly picked N anchor–anchor vectors from 4000 independent simulation trajectories, each containing 3000 time points and fitted the analytical model to these data. Because the covariance of $P(\vec{R}) = P_x(\delta x)P_y(\delta y)P_z(\delta z)$ is zero, rather than fitting this 3D function to the anchor–anchor vectors $\vec{R}$, we simultaneously fitted the three 1D probability densities $P_x(\delta x)$, $P_y(\delta y)$ and $P_z(\delta z)$ (Eq. 2) to the three axial projections of anchor–anchor vectors (δx,δy and δz), respectively. We performed these fits using a Python script that employs the curve_fit function of the scipy package (55), with the three proportions initialized to 1/3 each ($A_{\text{closed}} = A_{\text{open}} = A_{\text{extruding}} = \frac{1}{3}$). Note that to perform this fit we assumed the values of σ^2_{open} and σ^2_{closed} defined in section '*Anchor–anchor vector distributions*' above to be known (see Discussion).

Segmentation of time series into closed states. In the Results section 'Quantifying closed loop lifetimes from live-cell trajectories', we analyzed time series of anchor–anchor distances to estimate the duration of closed states. To do this, we segmented these time series into intervals of (inferred) closed states based on a spatial and a temporal threshold (Supplementary Figure S8A). Our procedure to detect closed states in simulated time series of anchor–anchor distances as function of time is as follows. First, we defined the spatial threshold as the 99.9% quantile of the anchor–anchor distances in the closed state (assumed to be known independently, see Discussion). The anchor–anchor distance in the closed state fluctuates around the diameter of 1 bead (50 nm). This is similar to the ∼40 nm diameter of the cohesin ring (56,57), which may enclose the two anchors in the closed state (17,26,58,59). Second, from time series obtained in simulations without extrusion, we measured the duration of time intervals during which the anchor–anchor distance was always below the spatial threshold (ignoring intervals reduced to a single time point). We then defined the temporal threshold as the 99.9% quantile of these time intervals. Within a time series, all time intervals with anchor–anchor distances below the spatial threshold and with durations exceeding this temporal threshold were segmented as closed state intervals. This resulted in binary time series with values of 1 for inferred closed states and 0 otherwise (i.e. for open or extruding states). In order to reduce the number of false negatives in closed state detections due to brief fluctuations of the distance above the spatial threshold, we then applied a rolling average to this binary time series (with a temporal window equal to the temporal threshold), and relabelled as closed states all timepoints with values above 0.5.

Estimation of the mean closed state lifetime. In the Results section 'Quantifying closed loop lifetimes from live-cell trajectories', we analyzed the segmented time series to estimate the mean lifetime of closed states assuming an exponential distribution of closed states (Figure 5C). This was done by fitting a 2-parameter linear function to the logarithm of the histogram of the durations of segmented closed state intervals. In order to reduce the influence of spurious closed states, we used a robust least-squares fit combined with a 'soft_L1' loss function from scipy (55), and the f_scale

parameter set to 0.002. We assessed the quality of the fit with 4000 bootstrapped samples of 1000 simulations each, drawn from a total of 10 000 independent simulations (Supplementary Figure S9A).

Estimation of extrusion speed. In the Results section 'Quantifying the speed of loop extrusion from live-cell trajectories', we estimated the extrusion speed by fitting an analytical function to $\langle\langle R^2\rangle\rangle(t)$, the ensemble average of the MSAAD (EMSAAD) over many time series, where $t = 0$ is the start of the closed state for all aligned time series. As described above (section '*Mean anchor–anchor distance as function of time*'), the theoretical EMSAAD follows a quadratic dependence with time, given by (Eq. 4). Nevertheless, for simplicity, and because the theoretical EMSAAD is close to a linear function for $t > -s_0/V_0$ (Supplementary Figure S10), we fitted a linear curve as function of time, whose slope defines an effective extrusion speed V_{eff}:

$$\langle\langle R^2\rangle\rangle(t) \approx R_0^2 + (V_{\text{eff}}/s_0)(-t) + R_{\text{loc}}^2 \quad (5)$$

After segmentation of closed state intervals (see section '*Segmentation of time series into closed states*' above), we defined a window size of 150 s, which is the minimal duration of extrusion for a 150 kb loop with V_0= 1 kb/s. In general, the minimal duration of extrusion is unknown, but it can be estimated in experimental data by choosing the window size with the highest estimated extrusion speed (see Results section 'Quantifying the speed of loop extrusion from live-cell trajectories'; Supplementary Figure S9B). We removed all intervals shorter than this time window. We then computed the EMSAAD as function of time and fitted a linear function using least_square from scipy (55) (Figure 5D, Supplementary Figure S9B–F).

Analysis of experimental data

Analysis of ChIP-Seq and Hi-C data. For all genomic analyses, we used the hg19 reference genome. For ChIP-Seq data of CTCF, SMC1 and RAD21, we used publicly available data from Rao *et al.* (16). Raw reads were quality-checked with FastQC (60). Reads from different replicates were first mapped independently using bowtie2 v2.2.6.2 (61) with default parameters, and the correlation between replicates was computed using wigCorrelate (62). Replicates with correlations larger than 0.9 were pooled together and mapped again. We removed blacklisted regions (63) and called peaks using MACS2 v2.1 (64) with default parameters. CTCF motifs were identified genome-wide using FIMO (65) with the flags -max-stored-scores 50 000 000 and –thresh 0.001. We then mapped CTCF sites identified with a P-value $< 1 \times 10^{-5}$ onto CTCF ChIP-Seq peaks.

For Hi-C data of HCT-116 cells, we used publicly available contact maps at 5 kb resolution from Rao et al. (16). To call contact frequency peaks genome-wide, we used HICCUPS from Juicer 1.19.02 (66) with the flags -r 5000, 10 000 -k KR -f 0.1 -p 4,2 -I 7,5 -t 0.02,1.5,1.75,2 and -d 20 000, 20 000. ChIP-Seq peaks of CTCF, SMC1 and RAD21 were intersected with 20 kb regions centered around the loop anchors using pgltools (67) intersect1D. For comparison with simulation predictions, we selected experimental loops with at least one pair of CTCF-bound convergent CTCF sites, SMC1 and RAD21 ChIP-Seq peaks at both anchors and genomic lengths ±10% of the simulated loop size.

From the experimental Hi-C maps at 5 kb resolution, we extracted the count matrices of selected loop domains using the dump function of Juicer 1.19.02 (66) (with the flags 'Observed' and 'NONE'). We then converted them to full matrices using the sparse2full function of HiCcompare (68). To identify contact frequency peaks in experimental Hi-C maps re-binned at 15 kb, we searched for the coordinates of the maximum intensity value in a 75 × 75 kb box centered on peak locations identified by Juicer (66). We removed experimental Hi-C maps that had a peak score above 5 since they were found to originate from artifacts (either bins with unusually low signal in the 30 kb × 30 kb background window or extremely high peak signal) and were outliers of the peak score distribution.

To build aggregated maps (Figure 2C), we extracted a 150 × 150 kb window around the loop anchors in each single Hi-C map. We summed these cropped Hi-C maps together and normalized by the number of maps used. We then binned the aggregated map to 15 kb resolution.

Analysis of imaging data. In the Results section 'Comparing polymer simulations to imaging data', we considered data from two experimental studies (Supplementary Figure S4). First, we used multiplexed DNA FISH data in mouse embryonic stem cells (mESC) from replicates 1 and 2 of Takei *et al.* (31). This dataset covered one genomic region on each of the 20 chromosomes at 25 kb resolution, each region comprising 60 consecutive loci and spanning at least 1.5 Mb, from which we extracted pairs of genomic loci. We removed from the analysis all cells where more than two identical chromosomes were detected. The pair-wise distances from all chromosomes were pooled together. Second, we considered live-cell imaging data from Gabriele *et al.* (35). From this study, we used data from the C36 clone without auxin treatment or with 4 h auxin treatment (which leads to RAD21 depletion).

From the simulations used for comparisons with these data, and for each considered genomic distance between pairs of loci, we computed the distributions of anchor–anchor distances assuming mean fractions of loop states consistent with experimental Hi-C data (as in Figure 2). We subsequently added random localization errors ($\sigma_{x,y}$ = 41 nm, σ_z = 65 nm) consistent with Takei *et al.* (31).

RESULTS

Overview of simulation and analysis workflow

Our goal is to examine, using simulations, if, how and to what extent loop extrusion can be detected and quantitatively characterized by imaging two fluorescently labeled loop anchors. We considered two types of experiments: (i) imaging fixed or living cells at a single time point (hereafter called static imaging), which provides a snapshot of anchor–anchor distances in a number of single cells and (ii) dynamic imaging of living cells, where the two anchors can be tracked and the anchor–anchor distance can be measured and followed as function of time (Figure 1A). Crucially, we assumed that extrusion can be experimentally abolished,

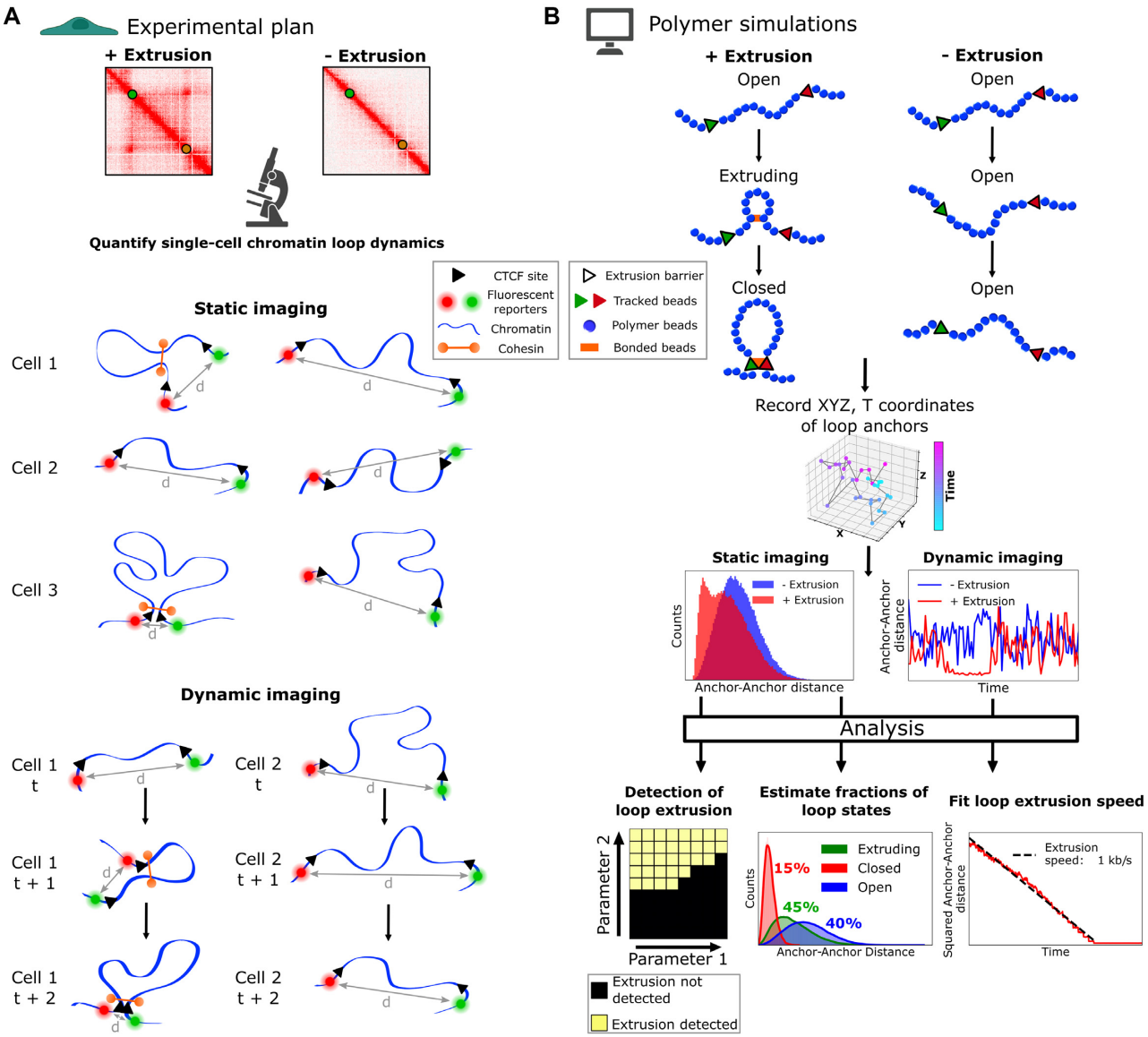


Figure 1. Overview of simulation and data analysis framework to characterize loop extrusion from imaging data. (**A**) Schematic of the assumed experimental plan. We assumed that 3D imaging data are obtained from cells in which extrusion can be abolished, e.g. by depleting cohesin. The two anchors of a loop are fluorescently labeled in two different colors (here, red and green) and imaged either at a fixed time point (static imaging) or tracked using time-lapse microscopy (dynamic imaging). This allows to determine the approximate 3D coordinates of the two anchors and the anchor–anchor distances (d) in a population of cells, either as a single snapshot or as function of time in single cells. Hi-C data shown on top are from Rao *et al.* (16). (**B**) Polymer simulations were used to model the dynamics of a generic chromatin segment with or without loop extrusion. The chromatin loop was assumed to transition between three different states, in the following order: open (the polymer is free of loops), extruding (the loop is actively extended) and closed (the loop is fully formed and the two anchors are kept close to each other). We used these simulations to generate 3D coordinates of each anchor as function of time (trajectories in the middle). From these trajectories, we obtained simulated distributions of anchor–anchor distances in cell population snapshots (static imaging) or simulated time series of anchor–anchor distances (dynamic imaging), either in absence (blue) or presence (red) of loop extrusion. We then used dedicated analysis methods to detect the presence of extrusion (bottom left), to estimate the fractions of cells in each of the three loop states (bottom center), or to determine the closed state lifetime and the speed of extrusion (bottom right). Based on the simulations, we determined the ranges of biological and experimental parameters (here noted generically parameters 1 and 2) for which extrusion can or cannot be detected (yellow and black, respectively) or for which loop state fractions or extrusion speed can be determined.

e.g. using an auxin-dependent degron (69,70) fused to RAD21, thus providing an essential negative control without loops in the genome (16) (Figure 1A).

Our analysis comprises the following steps, sketched in Figure 1B. First, we simulated the dynamics of polymers in presence or absence of loop extrusion and assuming various biological parameters, such as the loop size or the speed of loop extrusion. We supposed that the loop can be found in three different states hereafter called 'open', 'extruding' and 'closed'. In the open state, the chromatin segment delimited by the two anchors is free of any loop, e.g. because no extrusion complex is bound to it; in the extruding state, a loop located in between the two anchors is progressively enlarged by the action of an extrusion complex; in the closed

state, the two anchors are brought in close proximity and the loop is temporarily stabilized (the extrusion complex maintains the loop anchors together). Second, we compared simulated contact maps to available Hi-C data in order to determine plausible ranges for the fractions of loops in each state (open, extruding or closed) within a population of cells. Third, we used the same simulations to predict distributions of anchor–anchor distances for different biological and experimental parameters. We then used a statistical test to determine (for each set of parameters), whether and to what extent the presence of loop extrusion can be detected from distance distributions. Fourth, we proposed and tested a method to estimate the fractions of the open, closed, or extruding states based on distance distributions from static imaging data. Fifth, we analyzed how live-cell anchor trajectories can be used to estimate the duration (lifetime) of closed states and the speed of loop extrusion. In the following, we detail each of these steps and determine the conditions under which extrusion can be detected and quantified.

Simulating polymer dynamics and loop extrusion

In the following, we considered a generic, 1.8 Mb long chromatin segment centered around two loop anchors separated by 300 kb (unless otherwise stated). We modelled this chromatin segment as a polymer chain made of 600 beads linked by harmonic bonds and subjected to Langevin dynamics. We verified that mean squared displacements (MSD) of an internal bead in a non-extruding polymer (i.e. in the 'open' state) obeyed the subdiffusive power law expected from Rouse dynamics (MSD proportional to the square root of time) at small time scales and saturated at long time scales due to confinement (Supplementary Figure S1C).

To model loop extrusion, we assumed that a single loop extrusion complex was loaded on the polymer at a random position between the two anchors, whereafter both sides of the polymer chain were extruded simultaneously (bidirectional extrusion (26,71)). Unless stated otherwise, we assumed that extrusion began at t = 333 s after the start of the recorded simulation trajectories and proceeded at a nominal speed of $V_0 = 1$ kb/s (17). As soon as the extrusion complex reached one of the two anchors, it stopped, and the other side of the chain continued to be extruded unidirectionally (18,50,51), at a speed of $V_0/2$ (26), until the extrusion complex reached the second anchor (Supplementary Figure S2A). At this point, extrusion stopped entirely, and the two anchors were maintained in the closed state until the end of the simulation, unless stated otherwise. For any given set of biological or experimental parameters, we ran at least 1000 independent simulation trajectories, with ≈12 million time steps each. From these simulations, we directly extracted anchor coordinates as function of time, mimicking live-cell tracking data, or computed the distributions of anchor–anchor distances (and anchor–anchor vectors) in a random ensemble of configurations, mimicking population snapshots taken at a single time point.

Comparing polymer simulations to Hi-C data

Since TADs and loops were primarily defined from Hi-C data (4,5,16), realistic simulations must be able to account for these features in simulated contact frequency maps. In order to generate contact frequency maps from simulations, we used a capture radius of 150 nm, which provided the best match between simulated and experimental $P(s)$ curves (Supplementary Figure S2B). We first verified that our simulations, in which extrusion was assumed to start at random genomic locations between the anchors, can qualitatively reproduce blocks of enriched contacts with corner peaks on the contact map diagonal, consistent with TADs and loops (Supplementary Figure S2C). Note that if extrusion was instead initiated at one of the two anchors or at the mid-point between them, the contact maps exhibited different features, namely stripes (18,52,72), and hairpins (73) or jets (50), respectively (Supplementary Figure S2C). However, we excluded such structures from our analyses to focus on canonical TADs and loops instead.

The strength of contact blocks and/or corner peaks in Hi-C data is expected to depend on the fractions of cells in which the anchors are in an open, closed, or extruding state. Obviously, these fractions also strongly determine the dynamics of tracked anchors and the expected distribution of anchor–anchor distances, and therefore impact the feasibility of detecting and/or characterizing loop extrusion (37). For example, a high fraction of cells in a closed state could give rise to a bimodal distribution of anchor–anchor distances (one mode at small distances corresponding to closed states and a second mode at larger distances corresponding to open or extruding states), whereas a low fraction might not. These fractions are poorly known - one aim of imaging-based approaches is precisely to estimate them. Nonetheless, to determine a realistic range of state fractions, we proceeded to vary them in our polymer simulations and compared simulated contact maps to a selection of experimental loops in Hi-C data (Figure 2, Supplementary Figures S2E, F and S3). Below, we describe (i) the Hi-C data set considered, (ii) the simulations and the range of parameters explored, (iii) the metric we used to compare data to simulation predictions and (iv) the results of this comparison.

We started from the Hi-C data (16) of RAD21-mAID-GFP tagged HCT-116 cells, in which addition of auxin depletes RAD21 (69). We computationally identified 4470 loops in the 5 kb resolution Hi-C maps of cells without auxin treatment. We then narrowed down this data set to 229 loops that satisfied the following criteria: (i) a size of 300 ± 30 kb, for consistency with the 300 kb size of simulated loops, (ii) the presence of ChIP-Seq peaks of the cohesin subunits SMC1 and RAD21, as well as ChIP-Seq peaks of CTCF at both loop anchors, and (iii) at least one pair of bound convergent CTCF sites at the anchors (Figure 2A). These selection criteria were intended to focus the analysis on loops that depend on cohesin-mediated loop extrusion stabilized by CTCF, rather than on other biological mechanisms (10,74).

Using 200 000 single polymer conformation snapshots from 2500 independent simulations, we built 231 distinct contact maps, each corresponding to a different combination of the open, extruding or closed state fractions (Supplementary Figure S3). The extreme cases, where all conformation snapshots are from the same state, showed the expected behavior: for 100% open states, the contact map displayed no specific structure, and only reflected the uniform decay

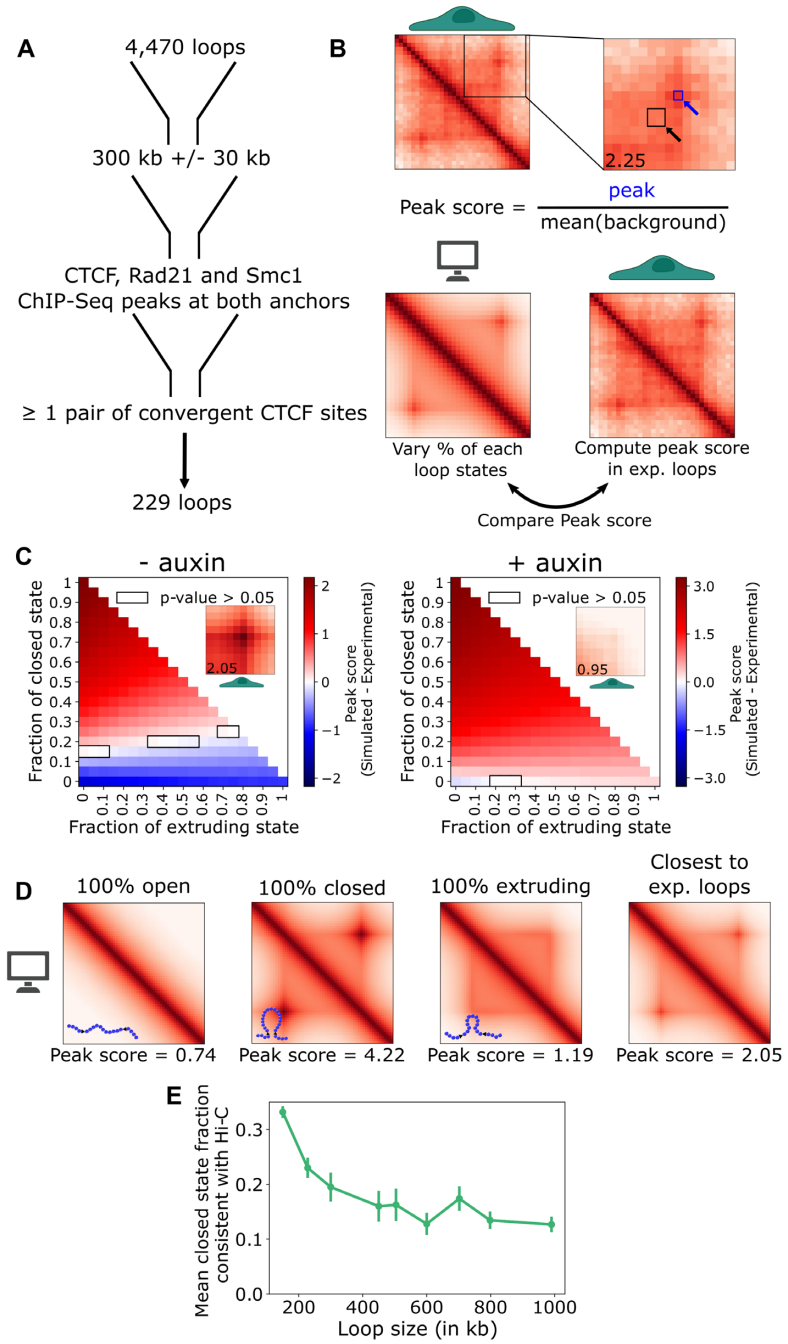


Figure 2. Estimation of loop state fractions consistent with experimental Hi-C data. (**A**) Starting from 4,470 loops detected in Hi-C data of HCT-116 cells from Rao et al (16), we selected a subset of 229 loops with sizes of 300 ± 30 kb, with ChIP-seq peaks of CTCF, RAD21 and SMC1 at both anchors and at least 1 pair of CTCF-bound convergent CTCF sites. (**B**) Top: For each of these loops, a peak score was computed as the ratio of the peak Hi-C contact frequency between the two anchors (blue square) and the background contact frequency (average of the black square). Bottom: Peak scores were used to compare simulated contact frequency maps (for varying fractions of each loop state) to experimental Hi-C maps. (**C**) Heat maps show differences in peak scores between simulated and experimental contact maps for each combination of extruding, closed and open state fractions, for 300 ± 30 kb loops. The cells were subjected or not to a 6 h auxin treatment to deplete the cohesin subunit RAD21 (16). Red and blue colors indicate a positive or negative difference between the simulated and the experimental mean peak scores, respectively, whereas white indicates similar peak scores. The open state fraction is: $A_{\text{open}} = 1 - A_{\text{closed}} - A_{\text{extruding}}$ where A_{closed} and $A_{\text{extruding}}$ are the fractions of closed and extruding states, respectively. The black contour denotes state fractions for which the peak scores of simulated and experimental loops were indistinguishable (P-value > 0.05 for a one sample t-test). Insets show aggregated maps of the experimental Hi-C data centered on the loop anchors and their mean peak score is displayed on the bottom left. The same display range was used for all aggregated maps. (**D**) Simulated contact maps for selected fractions of loop states. The first three maps from the left show extreme cases corresponding to 100% open, 100% closed or 100% extruding states. The fourth map corresponds to the state fractions that achieved the best match with the experimental Hi-C data (40% open, 40% extruding and 20% closed). The peak score for each simulated map is indicated below. All contact maps are shown with the same display range. Simulated and experimental Hi-C maps in panels B–D are shown at 15 kb resolution. (**E**) Mean fraction of closed states in simulations that are consistent with Hi-C data (P-values > 0.05) for loops of different genomic sizes. Error bars indicate the 95% confidence interval.

of contact frequencies as function of genomic distance, $P(s)$; for 100% closed states, the contact map exhibited a strong peak at the corner of a moderately intense contact block; for 100% extruding states, the map showed a block of enhanced contact frequencies delimited by the anchors, without a corner peak (Figure 2D and Supplementary Figure S3).

For a more quantitative comparison of simulation predictions with Hi-C data, we computed a 'peak score' in individual contact maps. The peak score was defined as the contact frequency between the two anchors divided by the contact frequency inside a 30 × 30 kb window located between the two anchors (Figure 2B). A higher fraction of closed states led to a high peak score, while a higher fraction of extruding states led to a higher background and thus a lower peak score (Figure 2D). We then systematically compared simulated and experimental peak scores for fractions of open, extruding and closed states ranging from 0 to 100%. For this purpose, we first verified the consistency of our metric with experimental findings that cohesin depletion leads to elimination of chromatin loops using Hi-C maps of auxin-treated cells (16). As expected, only simulations without closed states and with a large majority of open states (>70–80%) had peak scores statistically consistent with the Hi-C data in these cells (Figure 2C, right; $P > 0.05$).

We then used this peak score to compare simulation predictions to Hi-C maps in cells without auxin treatment. We found that simulation predictions were consistent with the Hi-C data for a range of closed state fractions of 15–25% and a large range of extruding state fractions (Figure 2C, left; $P > 0.05$). Within these ranges, the best match between simulations and Hi-C data was obtained for 20% closed states, 40% extruding states and 40% open loop states (Figure 2C, D). Simulations with 100% closed states were inconsistent with the experiments, leading to peak scores two times higher than in the Hi-C data (Figure 2D, $P < 10^{-150}$), consistent with previous studies arguing against the existence of stable loops in cells (20,33,75,76). However, our metric did not enable us to determine tight constraints on the fractions of open versus extruding states, as a wide range of fractions for each of these two states was compatible with the Hi-C data (Figure 2C and Supplementary Figure S2E).

To account for differences in the genomic sizes of loops, we then repeated the above simulations and analyses for loop sizes ranging from 150 to 990 kb. We found that the mean fraction of closed states consistent with the Hi-C data was larger for small loops than for large loops (33% for 150 kb loops, versus 16% for 504 kb loops and 13% for 990 kb loops) (Figure 2E and Supplementary S2F).

Thus, we used Hi-C data to estimate plausible ranges for the fractions of open, closed and extruding states in our polymer simulations. This allowed us to subsequently simulate loop extrusion dynamics and imaging observations under realistic conditions.

Comparing polymer simulations to imaging data

Having determined the fractions of loop states consistent with experimental Hi-C data, we next compared our simulations to published measurements of distances between chromatin loci from experimental imaging data. We first compared simulations of 504 kb loops with or without extrusion to recent live-cell imaging data from a study by Gabriele *et al.* (35) that specifically targeted a pair of loop anchors in mouse embryonic stem cells (mESC) with and without cohesin and found good agreement between experiments and simulations in both cases (Supplementary Figure S4). Next, we compared our simulations to a larger experimental data set where DNA FISH was used to measure distances between pairs of loci in mESC for a large range of genomic distances between these loci (31). These experiments did not specifically target loop anchors, but loops were nonetheless present in the targeted genomic regions. As a result, we expected anchor–anchor distances predicted by our simulations with and without extrusion to bracket the distances measured by DNA-FISH. This was indeed the case, and distances predicted by simulations with extrusion were slightly lower than the experimentally measured distances and closely followed the predicted trend as function of genomic distance (Supplementary Figure S4). These comparisons indicated that our simulations were also consistent with imaging data in fixed and living cells.

Detecting chromatin loop extrusion from imaging data

Chromatin loop extrusion is expected to reduce anchor–anchor distances compared to conditions without extrusion. We therefore asked whether loop extrusion can be detected simply by measuring anchor–anchor distances in populations of single cells and comparing them to distances measured when loop extrusion is abolished (Figure 3A). We aimed to assess the ability to reveal loop extrusion (hereafter called 'detectability' and defined quantitatively below) depending on several relevant biophysical and experimental parameters. The biological parameters were: the fractions of loop states (open vs. extruding vs. closed), the genomic size of loops, and the speed of loop extrusion. The experimental parameters were: the reporter-anchor separation, which we varied from 0 to 300 kb, the localization precision of fluorescent labels, defined as the standard deviation of errors in estimated coordinates, which we varied from $\sigma_{x,y} = 0$ nm to $\sigma_{x,y} = 200$ nm laterally (with an axial error $\sigma_z = 2\sigma_{x,y}$ due to the anisotropy of the microscope point spread function (54)) and the number of analyzed anchor–anchor distances, which we varied from $N =$ 10 to 10 000.

In order to determine whether extrusion can be detected, we compared the distributions of anchor–anchor distances in simulations with and without extrusion, thereby mimicking the experimental comparison of cells with and without cohesin. For each pair of simulated distance distributions, we performed Kolmogorov–Smirnov tests on thousands of random samples. We then defined the detectability of loop extrusion as the fraction of tests that revealed statistically significant differences between these distributions (Figure 3B).

To explore the effect of the different parameters above on the detectability of extrusion (Figure 3C–I), we varied two parameters at a time, holding the other parameters constant. We used the following state fractions as defaults, based on the above comparison to Hi-C data: 20% closed, 40% extruding, 40% open (Figure 2C). Other

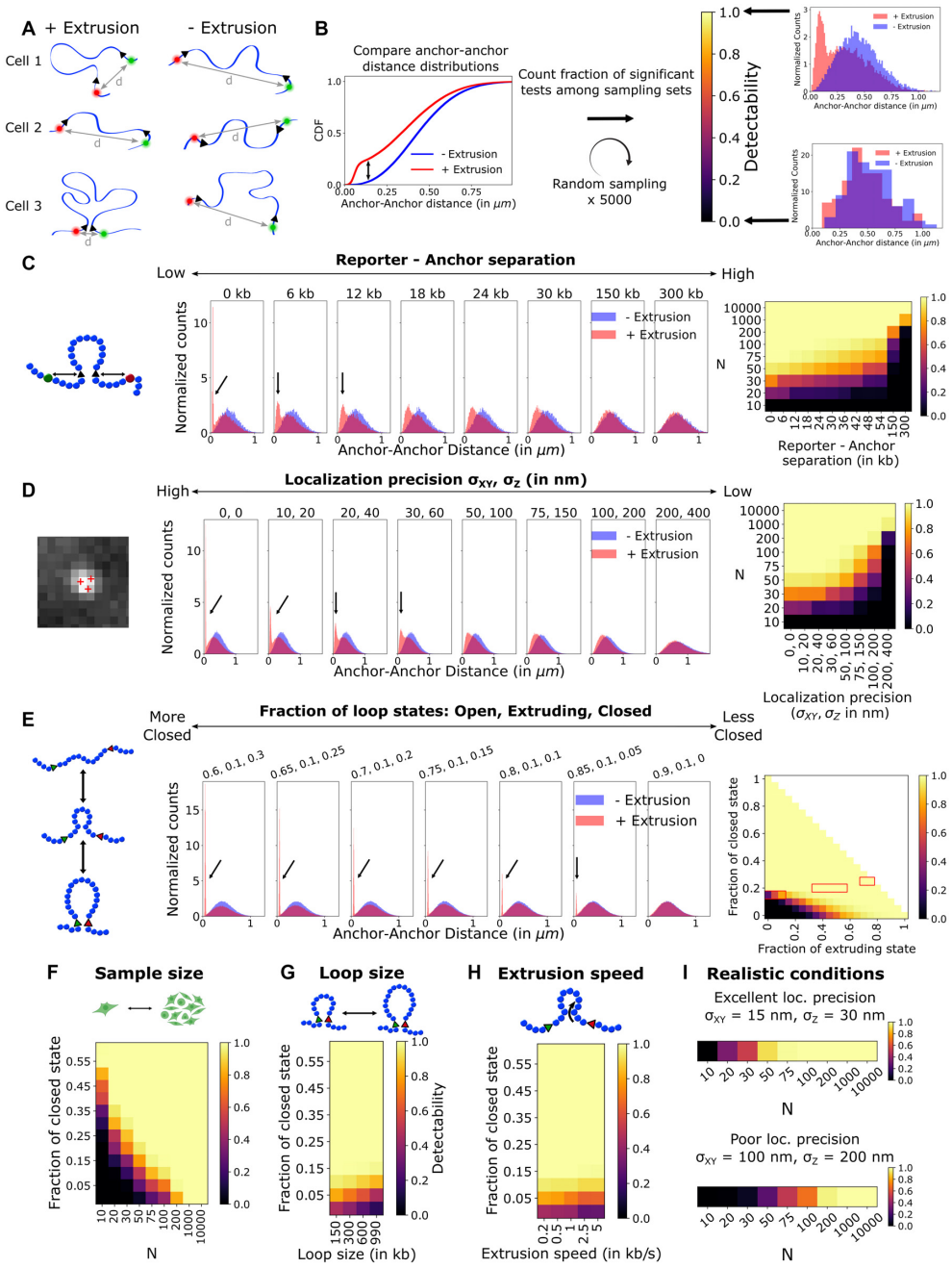


Figure 3. Experimental and biological parameters allowing to detect loop extrusion. (**A**) Polymer simulations assumed that distances between fluorescently labeled anchors are measured in a population of cells at a given time, both in conditions allowing extrusion, and in a condition where extrusion is experimentally abolished. (**B**) Cumulative distribution functions (CDFs) of anchor–anchor distances sampled from simulations with (red) or without (blue) extrusion were compared using a Kolmogorov-Smirnov test. Sampling was repeated 5000 times, *P*-values were corrected for multiple testing with the Benjamini–Hochberg procedure. Detectability of loop extrusion was defined as the fraction of tests revealing significant differences ($P < 0.05$) and is displayed on a color scale from black (no significant difference) to yellow (all tests show significant differences). Histograms show examples of distance distributions for these extreme cases of low and high detectability. (C–H) Effect of experimental or biological parameters on the detectability of loop extrusion. Heat maps show detectability as function of various parameters, including: the reporter-anchor separation (**C**), the localization precision of fluorescent reporters $\sigma_{x,y} = \sigma_z/2$ (**D**), the fractions of closed and extruding states (**E**), the sample size (i.e. the number of distance measurements), N (**F**), the genomic size of the loop (**G**) and the speed of loop extrusion (**H**). Two parameters were varied at a time, with the other parameters set to the following default values: fractions of loop states: 40% open, 40% extruding, 20% closed; loop size: 300 kb; extrusion speed: 1 kb/s; reporter-anchor separation: 0 kb; localization precision: $\sigma_{x,y} = \sigma_z = 0$ nm; sample size: $N = 100$. Distributions of anchor–anchor distances with and without extrusion (red and blue, respectively) and for different parameter values are shown in (C)–(E) (note that we used $N = 100\,000$ to plot these distributions for better legibility, irrespective of the sample size N used for the detectability heat maps). The arrows show the peaks at small distances due to closed states. Note that 'anchor–anchor distance' designates the distance between the corresponding reporters (which differ from the actual anchors for panel C). In the heat map of panel (E), the red contour denotes loop state fractions consistent with experimental Hi-C data from Figure 2C (without auxin treatment). (**I**) Detectability of loop extrusion in realistic experimental conditions with: reporter-anchor separation of $2 \times 3 = 6$ kb and excellent ($\sigma_{x,y} = \sigma_z/2 = 15$ nm, top) or poor ($\sigma_{x,y} = \sigma_z/2 = 100$ nm, bottom) localization precision. Other parameters were set to the default values above.

default parameters were as follows: loop size: 300 kb; extrusion speed: 1 kb/s, based on measurements of DNA loop extrusion *in vitro* (17,26); localization precision $\sigma_{x,y} = \sigma_z = 0$ nm; reporter-anchor separation = 0 kb (corresponding to a direct labeling of the anchors); sample size: $N = 100$ measurements (anchor–anchor distances).

We first analyzed the effect of the reporter-anchor separation (Figure 3C). In the ideal case where this separation is zero (i.e. the anchors themselves are fluorescently labeled), the clearest signature of extrusion was an additional peak in the anchor–anchor distance distribution at very low distances due to closed states (Figure 3C, arrows). In this case, extrusion detectability was high (95%) for a sample size N as low as 50. As expected, increasing the reporter-anchor separation tended to reduce and broaden the low distance peak, and for a total separation of 18 kb or more the distribution became unimodal again (Figure 3C). Consistently, increasing the reporter-anchor separation lowered detectability when the sample size N was kept constant. However, even for reporter-anchor separations of 54 kb, extrusion remained detectable using $N > 100$ measurements (detectability>92%). Increasing the random localization errors had a qualitatively similar effect as increasing the reporter-anchor separation (Figure 3D). Nevertheless, even for poor localization precisions of $\sigma_{x,y} = \frac{\sigma_z}{2} = 100$ nm, detectability was high (>97%) as long as the sample size N exceeded 200 (Figure 3D).

Obviously, these results strongly depended on the assumed fractions of closed, extruding and open states. For example, for a fixed extruding state fraction of 10%, it became more challenging to detect extrusion from $N = 100$ measurements when closed state fractions were lower than 20% (Figure 3E). Interestingly, increasing the sample size to $N = 1000$ measurements allowed extrusion to be detected for almost all combinations of state fractions, except in the quasi-absence of closed states combined with an extremely high fraction of open states (>80%) (Supplementary Figure S5A).

We then assessed the effects of sample size, loop size and extrusion speed, assuming 40% of extruding states as based on the above analysis of Hi-C data. A larger sample size always helped to detect extrusion in unfavorable conditions. For example, for a small closed state fraction of 5%, increasing the sample size from $N = 100$ to $N = 1000$ improved detectability from 70% to 100% (Figure 3F). The loop size, which we varied between 150 and 900 kb, had only a marginal effect on detectability (Figure 3G). Similarly, varying the speed of extrusion in the range 0.2–5 kb/s, which encompasses extrusion speeds measured *in vitro* (17,26), did not appreciably affect the detectability of loop extrusion (Figure 3H).

Next, we considered the detectability of extrusion as function of localization precision and sample size for realistic values of all other parameters combined (Figure 3I). Specifically, we assumed 20% closed, 40% extruding and 40% open states; a loop size of 300 kb; an extrusion speed of 1 kb/s; and a total reporter-anchor separation of 6 kb. We tested excellent, good and poor localization precisions ($\sigma_{x,y} = \frac{\sigma_z}{2} = 15$, 50 and 100 nm (31,32,77), respectively). The presence of closed loops was identifiable as a low distance peak in the distance histogram for excellent localization precision, but not for good or poor precision (Supplementary Figure S5B). Nevertheless, even for poor precision, loop extrusion detectability was still high (96%) for a reasonable sample size of $N = 200$ (Figure 3I).

Finally, we performed simulations using experimental and biological parameters approximating those of the above-mentioned Gabriele *et al.* study (35), which reported the identification of loop extrusion from anchor–anchor distances measured using live-cell imaging. Based on these simulations, we found loop extrusion to be detectable at 97% with as little as $N = 100$ measurements, while at least 45 000 distances were measured in (35) (Supplementary Figure S6). Hence, this study's experimental approach indeed matches the conditions defined by our theoretical analysis to detect loop extrusion from imaging data.

In summary, our analysis suggests that under a relatively wide range of realistic biological and experimental conditions, the presence of loop extrusion can be detected in cells by imaging pairs of loop anchors, measuring their distances and comparing them to distances measured in cells where extrusion is abolished. The above results also define the minimal experimental parameters that are required (and satisfied by Gabriele *et al.* (35)) to detect the presence of loop extrusion depending on the fractions of different loop states.

Estimating the fraction of loop states from static imaging data

Next, we aimed to move beyond the mere detection of loop extrusion and assess if static imaging data from a population of single cells allows to quantitatively estimate the fractions of loop states (closed vs. extruding vs. open) (Figure 4). We approached this by fitting to the data a 'three-state' mathematical polymer model that includes these fractions as adjustable parameters. Specifically, we considered the differences in x, y and z coordinates between the two anchors ($\delta x = x_1 - x_2$; $\delta y = y_1 - y$; $\delta z = z_1 - z_2$) and fitted a model, $P(\delta x, \delta y, \delta z) = P_x(\delta x) P_y(\delta y) P_z(\delta z)$ to the three distributions of coordinate differences ($\delta x, \delta y, \delta z$) simultaneously. For each coordinate difference $\delta w \in \{\delta x, \delta y, \delta z\}$, $P_w(\delta w)$ is a linear combination of three models, P_w^{open}, P_w^{closed} and $P_w^{\text{extruding}}$, one for each loop state (open, closed and extruding, respectively), weighted by the fractions of each state (A_{open}, A_{closed}, and $A_{\text{extruding}} = 1 - A_{\text{closed}} - A_{\text{open}}$), and is given by (Eq. 2). For the open state, we assumed that the polymer behaves as a freely jointed chain and each coordinate difference $\delta w \in \{\delta x, \delta y, \delta z\}$ follows a normal distribution $P_w^{\text{open}} = \mathcal{N}(0, \frac{N_0 b^2}{3} + 2\sigma_w^2)$, where b is the Kuhn length, N_0 is the number of Kuhn lengths between the anchors, and σ_w is the localization precision along axis $w \in \{x, y, z\}$. For the closed state, where the two anchors are in contact or in close proximity, we assumed that each coordinate difference follows a normal distribution $P_w^{\text{closed}}(\delta w) = \mathcal{N}(0, \sigma_c^2 + 2\sigma_w^2)$, where the added variance σ_c^2 reflects a possible non-zero distance between the anchors in the closed state (e.g. because of the ~40 nm size of the cohesin ring (56,57)). We assumed that the probability distributions P_w^{open} and P_w^{closed} can be determined experimentally (see Discussion).

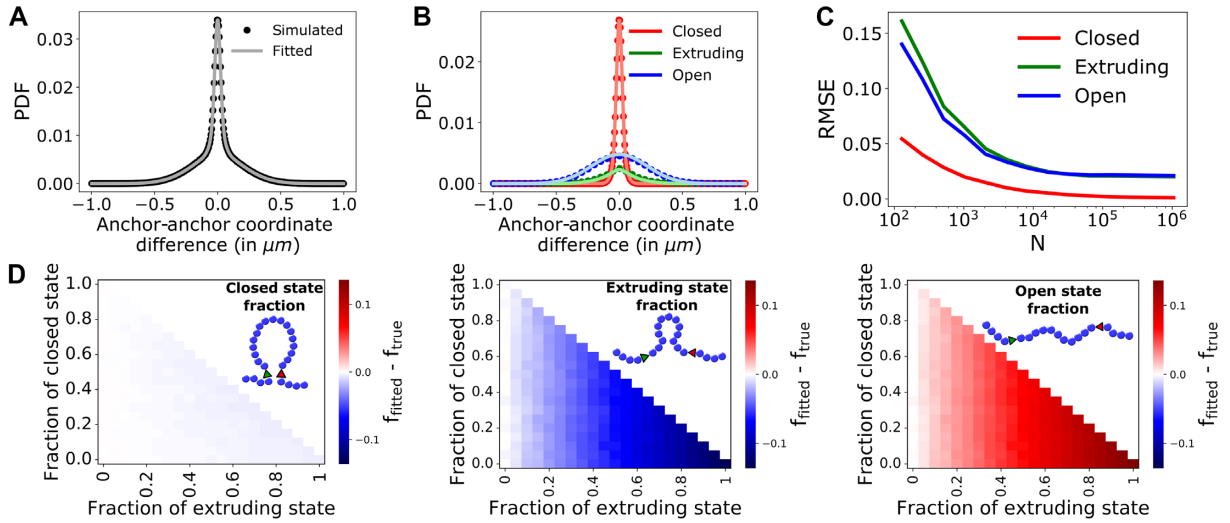


Figure 4. Quantifying fractions of loop states from static images. (**A**) Total distribution of the differences δw between coordinates of the two anchors ($w \in \{x,\ y,\ z\}$) as obtained from simulated data (black dots) and the fitted three-state model (grey curve, Eq. 2). (**B**) Same as A, but with the three fitted components of the three-state model (corresponding to the closed, extruding and open states) shown separately. Dots show the simulated data and colored curves show the fitted model components. (**C, D**) Quantification of errors in estimation of loop state fractions. (C) Root mean square error (RMSE) of estimated fractions for the three states as function of the sample size (number of measured anchor–anchor vectors) N. The simulations used for panels (A), (B) and (C) assumed the following loop state fractions: 35% closed, 15% extruding and 50% open. (**D**) Heat maps show the signed error (bias) in the estimated fractions of closed, extruding or open states, for each combination of loop state fractions, estimated for $N = 500\,000$. Overestimates are shown in red, underestimates in blue. The absolute bias did not exceed 13.1%. All simulations used in this Figure assume a localization precision $\sigma_{x,y} = \sigma_z/2 = 15$ nm.

For the extruding state, we assumed that the coordinate differences follow the probability density $P_w^{\text{extruding}}(\delta w; t) = \mathcal{N}(0, \frac{N_b(t)b^2}{3} + 2\sigma_w^2)$, with $N_b(t) = N_0 - N_{\text{loop}}(t)$, where $t = 0$ is the end of loop extrusion (during extrusion $t < 0$), and $N_{\text{loop}}(t)$ is the number of Kuhn lengths extruded at time t. This assumption effectively derives from the hypotheses that (i) the anchors in the extruding state behave as if part of a shorter polymer from which the extruded loop is absent and (ii) this polymer has time to equilibrate at each step of the extrusion process. In the Materials and Methods section, we derived an explicit model for $P_w^{\text{extruding}}(\delta w; t)$ (Eq. 1) based on the additional assumptions that extrusion initiates with uniformly random probability between the two anchors and switches from bidirectional at constant speed V_0 to unidirectional at constant speed $\frac{V_0}{2}$ upon reaching an anchor. Having specified these models for P_w^{open}, P_w^{closed} and $P_w^{\text{extruding}}$, we could estimate the fractions of each loop state (open, closed and extruding) by fitting the three-state model of (Eq. 2) to the measured distributions of coordinate differences δx, δy and δz.

To test this method, we first simulated coordinates of anchor pairs from 4000 independent simulations assuming 35% closed, 15% extruding and 50% open states (based on the analysis of Hi-C data in Supplementary Figure S2F) for a 150 kb loop, with localization errors $\sigma_{x,y} = \sigma_z/2 =$ 15 nm (29,31,32), drew N = 500 000 samples (anchor–anchor vectors) and fitted our three-state model (Eq. 2) to these data (Figure 4A-C). As before, these simulations assumed that extrusion initiates with uniformly random probability between the two anchors, and that the extrusion speed is halved once the first anchor is reached. As shown in Figure 4A and B, the fitted model was in very good agreement with the simulated data. We computed a root mean square error (RMSE) of estimated fractions A_{closed}, A_{open} and $A_{\text{extruding}}$ for varying sample size N (Figure 4C). For $N = 100$, the RMSE for open and extruding states remained substantial (14% and 16%, respectively), implying that larger sample sizes are needed for accurate estimation of loop state fractions. For $N > 5000$, the RMSE dropped to <3% for all three states, and decreased further for larger N. Open and extruding states still had RMSE <2.5% even for large sample sizes (Figure 4C), a residual bias that presumably reflects a minor mismatch between the analytical model and the simulations. We then systematically varied the fractions of open, extruding and closed states and quantified the absolute biases of the estimated fractions for a sample size of $N = 500\,000$ (Figure 4D). The closed state fraction was always very accurately estimated, with absolute bias <0.5%. The extruding state tended to be slightly underestimated, while the open state was slightly overestimated, but the absolute bias always remained at an acceptable level (<13.1%) (Figure 4D).

We repeated this analysis but changed the localization errors to either $\sigma_{x,y} = \sigma_z = 0$ or $\sigma_{x,y} = \sigma_z/2 = 50$ nm (Supplementary Figure S7). Paradoxically, the state fractions were estimated more accurately for larger localization errors (maximum absolute bias of 7.4%) than without localization errors (maximum absolute bias of 17.7%) (Supplementary Figure S7D and H), likely because the Gaussian localization errors reduce the mismatch between the Gaussian model and the simulations in the closed state. Nevertheless, for sample sizes $N > 8000$, the RMSE were <5% both without localization errors and for large localization errors, assuming 35% closed, 15% extruding and 50% open states (Supplementary Figure S7C and G).

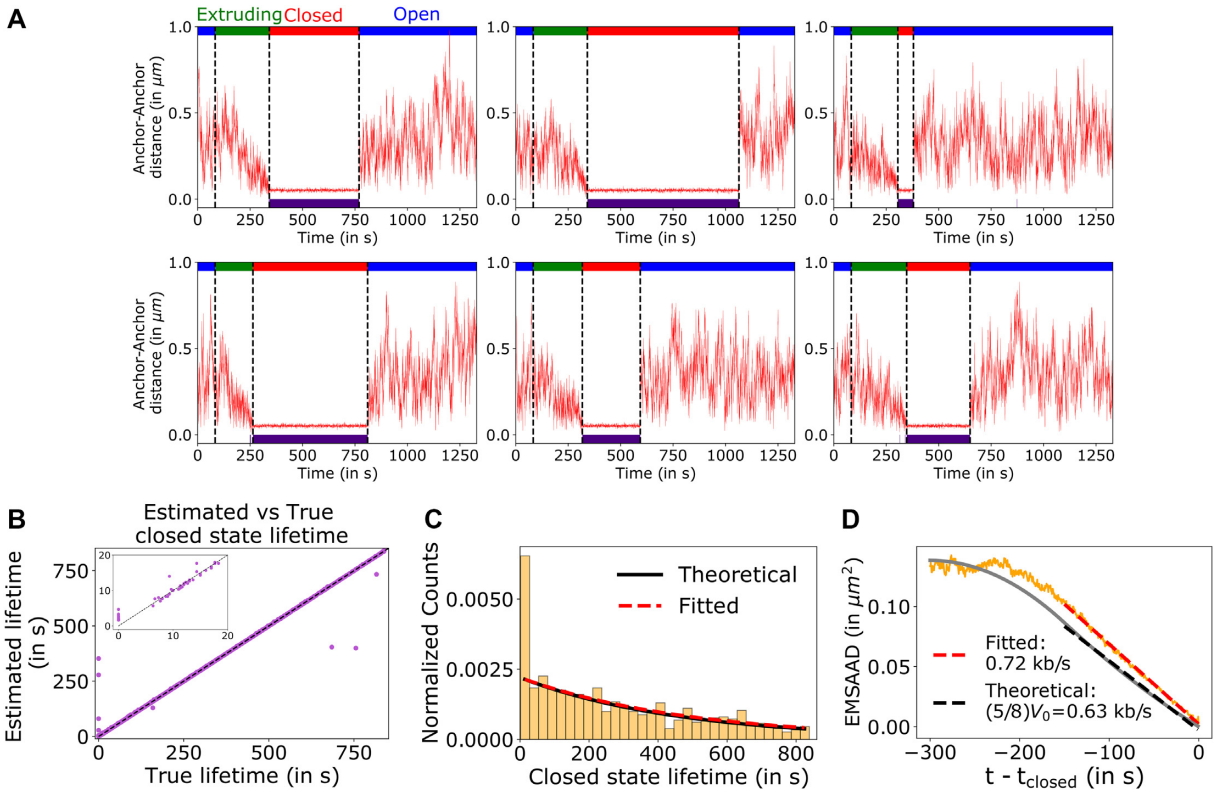


Figure 5. Quantifying the lifetime of closed states and the speed of loop extrusion by dynamic tracking of loop anchors. (**A**) Six example time series of anchor–anchor distances from simulations with loop extrusion. The true loop state is indicated by color bars on top (blue: open, green: extruding, red: closed). Black dotted lines indicate state changes, i.e. when loop extrusion starts or ends or when closed loops are released. The segmentation of the time series in intervals of closed states is shown below (indigo bar indicates inferred closed state intervals). (**B**) Estimated vs. true lifetimes of closed loops. Each indigo dot corresponds to a distinct segmented closed state interval. Black dotted line denotes $y = x$. The inset shows a close-up for short lifetimes (0–20 s). (**C**) Histogram of 1000 estimated lifetimes of closed states. The dotted red line shows a robust exponential fit to the segmented closed state lifetimes and the solid black line shows the theoretical distribution. (**D**) Ensemble mean squared anchor–anchor distance (EMSAAD) as function of time, averaged over >1000 time series, after aligning the estimated start time of the closed state to $t = 0$ (solid orange curve). An effective loop extrusion speed $\widehat{V_{\text{eff}}} = 0.72$ kb/s was estimated by fitting a linear function for $-150\text{ s} < t < 0$ (red dotted line). The theoretically expected EMSAAD (Eq. 4) is a non-linear function of time and is shown in solid grey. A linear fit to the theoretical EMSAAD (Eq. 5) is shown as a dashed black line and corresponds to a theoretical speed of 0.63 kb/s. Localization errors were ignored in the simulations used here.

To sum up, our analysis indicates that the fractions of loop states can be estimated accurately by measuring anchor–anchor distances from static imaging data of a sufficiently large number of cells.

Quantifying closed loop lifetimes from live-cell trajectories

While the analysis of static data can provide important quantitative information on loop states, as shown above, a full characterization of the dynamic aspects of loop extrusion, in particular extrusion speed and the lifetime of closed loops, requires dynamic data. Here and in the next section, we assess to what extent these parameters can be estimated from live-cell tracking of loop anchors.

For this purpose, we generated 1000 independent Langevin dynamics simulations of a polymer undergoing extrusion and recorded the trajectories of the two anchors as function of time. We considered a polymer with two loop anchors separated by 150 kb and simulated trajectories of 1330 s each. We assumed that extrusion started 83 s after the beginning of each simulation at a random position between the two anchors, and that extrusion proceeded until the loop was closed. The two anchors were then maintained in this closed state during a time interval drawn from an exponential distribution, in order to mimic a memory-less (Poisson) process of loop dissociation, with a mean closed state lifetime of 466.6 s (and a maximum of 833 s). After dissociation, the two anchors were released from each other, and the polymer relaxed rapidly to the open state. Typical time series of the anchor–anchor distance during a simulation trajectory are shown in Figure 5A.

In order to quantify closed state lifetimes, we aimed to segment these time series into disjoint time intervals corresponding to closed, open or extruding states. As apparent from the examples in Figure 5A, it is difficult to distinguish the open state from the extruding state, especially in the early phases of the extrusion process where the anchor–anchor distances fluctuate widely. We therefore focused on segmenting the closed state only, without attempting to distinguish open from extruding states. Nonetheless, the distance distribution in the closed state can still significantly overlap that of the extruding and open states, since in these two states anchors can still come into close vicinity due solely to random polymer movements (Supplementary Figure S8A). For this reason, a distance threshold alone is ill suited to segmenting closed states (Supplementary Figure

S8A). We therefore decided to use a combination of spatial and temporal thresholds and identified closed states as long-lived small distance states that were not present in time series without extrusion (as measured in an experimental condition that abolishes extrusion). We defined small distances as those below the 99.9% quantile of distances in the closed state. We defined the temporal threshold as the 99.9% quantile of time intervals during which the anchor–anchor distance was always below the spatial threshold in time series of extrusion-free simulations (Supplementary Figure S8A).

Next, we evaluated this segmentation method on simulated time series with extrusion by quantifying the number of time points correctly or incorrectly classified as closed states. Using the above spatial and temporal thresholds, closed states were identified with a precision (ratio of true positives over all positives) of 99.75% and a recall (ratio of true positives over the sum of true positives and false negatives) of 99.87%, proving high classification quality for individual time points (Supplementary Figure S8B). Precision and recall could be traded-off against each other by varying the spatial and temporal thresholds (Supplementary Figure S8B). We then analyzed the performance of our segmentation method for a total reporter-anchor separation of 6 kb and different localization errors. Recall remained above 99% for localization errors $\sigma_{x,y} = \sigma_z/2 < 50$ nm, while segmentation precision decreased when increasing localization errors (from 96% for $\sigma_{x,y} = \sigma_z = 0$ nm to 78% for $\sigma_{x,y} = \sigma_z/2 = 50$ nm) (Supplementary Figure S8C). Localization errors above $\sigma_{x,y} = \sigma_z/2 = 75$ nm drastically reduced both recall (from 93% for $\sigma_{x,y} = \sigma_z/2 = 75$ nm to 7% for $\sigma_{x,y} = \sigma_z/2 = 200$ nm) and precision (73% for $\sigma_{x,y} = \sigma_z/2 = 75$ nm vs. 49% for $\sigma_{x,y} = \sigma_z/2 = 200$ nm) (Supplementary Figure S8C).

We then proceeded to estimate closed loop lifetimes. This was much more challenging and error-prone, because any time point with a false negative closed state detection led to a fragmentation of segmented closed state intervals and therefore to an underestimation of the true lifetime, while false positive time points generally led to spurious short-lived closed states. Nevertheless, using the thresholds above, the estimated lifetimes of closed states were in excellent overall agreement with the true lifetimes (Figure 5B). The exception to this were very short lifetimes (Figure 5B, C), which originated from the residual false positives in the detection of closed states. 95% of these spurious closed states were detected during extruding states, typically at the end of the extrusion phase (Supplementary Figure S8D).

To estimate the mean lifetime of closed states, we fitted an exponential function to the histogram of estimated lifetimes using a robust approach that rejects outliers. We found that with this robust fit, we could accurately estimate the mean closed state lifetime (Figure 5C and Supplementary Figure S9A). The estimated mean closed lifetime across 4000 bootstrapped samples (each containing 1000 time series) was 465.9 s, in excellent agreement with the expected value of 466.6 s (Supplementary Figure S9A).

Therefore, our analysis suggests that closed state lifetimes can be accurately estimated by tracking loop anchors in live-cell imaging.

Quantifying the speed of loop extrusion from live-cell trajectories

Finally, we assessed whether loop extrusion speed could be quantified from 1000 time series of anchor–anchor distances obtained by live-cell imaging. Because the above approach identified closed loop intervals that depend on the presence of extrusion (thanks to the control condition that abolishes extrusion, Supplementary Figure S8A), we reasoned that timepoints before these intervals should be in the extruding state. Assuming, as before, that the polymer with a loop behaves at each time point as an ideal polymer from which the loop portion is absent, we derived analytical models for the temporal evolution of the mean squared anchor–anchor distance (MSAAD) $\langle R^2 \rangle$ (see Materials and Methods for details). If loop extrusion occurred at a constant speed $V(t) = V_0$, the MSAAD would follow a simple linear function of time, given by (Eq. 3). However, we assumed that the speed of extrusion switches from bidirectional with speed $V(t) = V_0$ to unidirectional with speed $V(t) = V_0/2$ when the extrusion complex reaches one of the two anchors (Supplementary Figure S2A). The time point of switching depends on the genomic site where extrusion is initiated: switching occurs earlier if this site is closer to one of the two anchors. As a result, fitting equation (Eq. 3) to measured MSAAD is expected to lead to an effective speed V_{eff} between $V_0/2$ and V_0, whose exact value depends on the site of initiation.

Under the additional assumption that the site of extrusion initiation (i.e. the loading site of the extrusion complex) is uniformly random between the two loop anchors, we further derived a model for the ensemble average of the MSAAD (EMSAAD) over many time series, denoted as $\langle\langle R^2 \rangle\rangle(t)$, obtained after aligning the starting times of closed states for all time series at $t = 0$, and given by Eq. (4) (Supplementary Figure S10B).

Equipped with this model, we proceeded to analyze time series simulated with a bidirectional extrusion speed $V_0 = 1$ kb/s that switched to a unidirectional extrusion with speed $V_0/2$ upon reaching an anchor. Although the dependence of $\langle\langle R^2 \rangle\rangle(t)$ on time in Eq. 4 is non-linear, we found that a linear fit of the EMSAAD, assuming a single speed V_{eff} (Eq. 5), led to a reasonably accurate estimation of the mean effective extrusion speed ($\widehat{V_{\text{eff}}} = 0.72$ kb/s estimated speed versus $(\frac{5}{8})V_0 = 0.63$ kb/s for the true mean effective speed, Figure 5D). The theoretical EMSAAD was well approximated by a linear fit of slope $-5/8\,R_0^2 V_0/s_0$ (Figure 5D). Therefore, multiplying the estimated effective speed $\widehat{V_{\text{eff}}}$ by 8/5 allowed us to recover the bidirectional extrusion speed V_0. This yielded $V_0 = 1$ kb/s for the theoretical curve, as expected, and $\widehat{V_0} = 1.15$ kb/s for the estimated speed. The remaining discrepancy between estimated and true bidirectional speeds ($\widehat{V_0}$ and V_0) can be attributed to two effects: (i) the simulated polymer does not have time to equilibrate given the assumed extrusion speed of $V_0 = 1$ kb/s, and (ii) the polymer is confined. Both effects are ignored by our model and lead to a departure from linearity of the EMSAAD as function of time. Indeed, the EMSAAD is closer to the theoretical (equilibrium) value in simulations with slower extrusion and a relaxed confinement (Supplementary Figure S11). Note that the estimated speed depends on

the size of the time window used for fitting. The optimal window size depends on the actual speed and is therefore not known, but correct time windows and speeds can be determined experimentally from a peak in the estimated speed as function of the time window, as shown in Supplementary Figure S9B. This is explained by the slope of the theoretical EMSAAD, which is maximum at $t = \frac{-s_0}{v_0}$ (Supplementary Figure S10B).

Finally, we assessed how experimental parameters affected the estimation of loop extrusion speed. Unsurprisingly, localization errors negatively impacted the estimation of extrusion speed (by deteriorating the accuracy of closed state segmentation). Localization errors of $\sigma_{x,y} = \frac{\sigma_z}{2} = 30$ nm led to underestimations of the speed by 7%, and $\sigma_{x,y} = \frac{\sigma_z}{2} = 75$ nm led to underestimations by 51% (Supplementary Figure S9C and S9E). Similarly, increasing the reporter-anchor separation from 0 to 30 kb led to an underestimate of the loop extrusion speed by 27% (Supplementary Figure S9D and F).

Nevertheless, this analysis suggests that under realistic conditions it should be possible to extract meaningful estimates of the speed of DNA loop extrusion by analyzing trajectories of loop anchors in time-lapse microscopy of living cells.

DISCUSSION

We used polymer simulations and presented dedicated analysis techniques to determine if, and under which conditions, chromatin loop extrusion can be detected and/or characterized quantitatively by imaging two loop anchors with fluorescence microscopy (Figure 1). We separately considered static imaging experiments, as can be performed in fixed cells, e.g. using DNA-FISH (28), or using single snapshots in live-cells, and dynamic imaging experiments in which two loop anchors are tracked over time (35,78). Our analyses assume that imaging experiments are performed both in cells where loop extrusion takes place and in cells where extrusion is experimentally abolished, e.g. by depletion of the cohesin subunit RAD21 (16), which provides a crucial negative control. We considered the effect of multiple biological and experimental parameters. Biological parameters include the fractions of closed, open and extruding states, the speed of loop extrusion and the genomic size of the loops. Experimental parameters include the reporter-anchor separation, the localization precision of fluorescent reporters, and the sample size (number of measured anchor–anchor distances).

To estimate plausible ranges of loop state fractions, we compared simulations to Hi-C data of ≈300 kb large loops (Figure 2), and found that only small (15–25%) fractions of closed states are consistent with the data, in line with previous studies (20,35). Despite this relative rarity of closed loops, we showed that in most cases the presence of loop extrusion can be detected from static images alone for realistic localization precisions, provided that the sample size N (i.e. the number of measured anchor–anchor distances) is large enough (Figure 3). For example, for loops of 300 kb with 20% closed and 40% extruding states, a reporter-anchor separation of 6 kb, and a lateral localization precision of 15 nm (30 nm axially), extrusion can be detected reliably from as little as $N \approx 75$ measurements (Figure 3I). With a poor lateral localization precision of 100 nm (200 nm axially), reliable detection requires $N \approx 200$ measurements, which is still within reach of standard microscopy experiments (Figure 3I). An important result of our analysis is that loop extrusion can be detected even in absence of a second peak in small anchor–anchor distances, which to our knowledge has never been observed in DNA-FISH (28,79) or live-cell imaging (35–38). Furthermore, we showed that analysis of static imaging data also enables to estimate the fraction of closed loops with high accuracy, and the fractions of loops in extruding and open states with reasonable accuracy (Figure 4). Finally, we demonstrated that under realistic conditions, dynamic imaging data can be analyzed to estimate the lifetime of closed states as well as the speed of loop extrusion (Figure 5). Our study therefore establishes the feasibility of an imaging approach to identify and quantitatively characterize loop extrusion. This is a non-trivial finding considering the many challenges arising from the highly stochastic nature of chromatin motion (80), the potential rarity of closed states, and technical limitations of image acquisition. We have defined the minimum conditions that are required to detect and characterize loop extrusion in terms of localization precision, proximity of reporters to anchors and sample size, and highlighted the critical importance of comparing cells with and without extrusion. We also verified that the recent experimental study of loop extrusion by Gabriele *et al.* in mouse embryonic stem cells indeed matched these conditions (35). The analysis methods proposed here (statistical comparisons of distance distributions and fitting to theoretical polymer models) should be directly applicable to experimental data in follow-up studies.

Nonetheless, we acknowledge that our analyses have caveats. First, our method for detecting loop extrusion (Figure 3) assumes that extrusion can be specifically abolished in experiments such as RAD21 depletion (16). We cannot exclude that potential changes to chromatin structure due to RAD21-depletion that are unrelated to extrusion, such as epigenetic modifications (81,82) or increased compartmentalization (23,24), may be picked up unspecifically by this approach. Second, in our analysis of dynamic imaging data, we identified the closed state as a temporally stable low-distance state absent from conditions without extrusion. This method will miss short-lived closed states whose duration is below the temporal threshold. The segmentation of closed states could potentially be made more specific by also considering conditions that permit loop extrusion but not closed states. Such conditions could be achieved experimentally by removing obstacles to loop extrusion, e.g. by depleting the CTCF protein or mutating CTCF binding sites (19,24). Third, our quantification of closed state fractions from static imaging data and our quantification of closed state lifetimes from dynamic data assume that the distributions of anchor–anchor distances in the closed state are known, which is not straightforward. One possible way to estimate these distances (together with localization errors) is to measure the distance between two fluorescent markers separated by a genomic distance equal to the summed distance between anchors and reporters. Fourth, we did not consider the possibility that multiple cohesin

complexes may simultaneously extrude chromatin loops between the two anchors (76,83,84), which would considerably complicate the analysis of state fractions and extrusion speed. Fifth, our analysis ignores the possibility that loops may be in the closed state at the start or end of dynamic imaging experiments (censoring). Sixth, our assumption of a continuous extrusion activity at constant (bidirectional or unidirectional) speed does not account for the possibility of pausing (85), which may lead to underestimations of the peak extrusion speed. Fully considering these complications will require follow-up work.

In spite of these limitations, our study highlights how imaging in fixed or living cells can be used to rigorously identify and quantitatively characterize the fundamental process of chromatin loop extrusion. With DNA-FISH, fractions of closed, extruding and open states could be estimated in a high-throughput manner by targeting probes against the anchors of TADs and loops for hundreds of genomic loci. Since each of these loci is associated with a unique epigenetic state, gene expression pattern and spatial neighborhood, such a systematic analysis might shed light on the intermingling between these factors and loop extrusion, and its consequences on genome functions (30,31,86,87). The possibility to analyze loop extrusion by live-cell imaging is illustrated by two above-mentioned experimental studies in mESC (35,36), which estimated closed loop fractions ranging from 3% to 27% and median loop lifetimes ranging from 5 to 30 minutes. We expect that subsequent experiments will extend this initial work by analyzing many more TADs and loops and their dynamics in living cells of multiple organisms. By providing clear quantitative guidelines and analytical techniques, we hope that the present study will be instrumental in reaching a better quantitative understanding of how loop extrusion affects chromatin dynamics and function.

DATA AVAILABILITY

The LAMMPS parameter files used for the molecular dynamics simulations and the Python code used to estimate loop state fractions from static imaging data are available at https://github.com/imodpasteur/LoopExtrusion and through Zenodo at (88). Numerical data of static and dynamic imaging simulations are available through Zenodo at (89). All other numerical data are available on request to the corresponding authors. Hi-C and ChIP-Seq data used in this work were taken from Rao *et al.* (16) with accession number GSE104334.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Jyotsana J. Parmar for stimulating discussions and initial help with the simulations. We thank Davide Normanno for critical reading of the manuscript. We acknowledge the help of the HPC Core Facility of the Institut Pasteur. Parts of some figures were created with BioRender.com.

FUNDING

Investissement d'Avenir [ANR-16-CONV-0005]; Contrat Doctoral Spécifique aux Normaliens; Fondation ARC pour la recherche sur le cancer; Institut Pasteur. Funding for open access charge: Institut Pasteur.
Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Parmar,J.J., Woringer,M. and Zimmer,C. (2019) How the genome folds: the biophysics of four-dimensional chromatin organization. *Annu. Rev. Biophys.*, **48**, 231–253.
3. Davidson,I.F. and Peters,J.-M. (2021) Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.*, **22**, 445–464.
4. 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.
5. 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.
6. Rajpurkar,A.R., Mateo,L.J., Murphy,S.E. and Boettiger,A.N. (2021) Deep learning connects DNA traces to transcription to reveal predictive features beyond enhancer–promoter contact. *Nat. Commun.*, **12**, 3423.
7. Espinola,S.M., Götz,M., Bellec,M., Messina,O., Fiche,J.-B., Houbron,C., Dejean,M., Reim,I., Cardozo Gizzi,A.M., Lagha,M. *et al.* (2021) Cis-regulatory chromatin loops arise before tads and gene activation, and are independent of cell fate during early Drosophila development. *Nat. Genet.*, **53**, 477–486.
8. Di Stefano,M., Stadhouders,R., Farabella,I., Castillo,D., Serra,F., Graf,T. and Marti-Renom,M.A. (2020) Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs. *Nat. Commun.*, **11**, 2564.
9. Greenwald,W.W., Li,H., Benaglio,P., Jakubosky,D., Matsui,H., Schmitt,A., Selvaraj,S., D'Antonio,M., D'Antonio-Chronowska,A., Smith,E.N. *et al.* (2019) Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. *Nat. Commun.*, **10**, 1054.
10. 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.
11. Zuin,J., Roth,G., Zhan,Y., Cramard,J., Redolfi,J., Piskadlo,E., Mach,P., Kryzhanovska,M., Tihanyi,G., Kohler,H. *et al.* (2022) Nonlinear control of transcription through enhancer–promoter interactions. *Nature*, **604**, 571–577.
12. Rodríguez-Carballo,E., Lopez-Delisle,L., Willemin,A., Beccari,L., Gitto,S., Mascrez,B. and Duboule,D. (2020) Chromatin topology and the timing of enhancer function at the HoxD locus. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 31231–31241.
13. Benabdallah,N.S., Williamson,I., Illingworth,R.S., Kane,L., Boyle,S., Sengupta,D., Grimes,G.R., Therizols,P. and Bickmore,W.A. (2019) Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell*, **76**, 473–484.
14. Xiao,J.Y., Hafner,A. and Boettiger,A.N. (2021) How subtle changes in 3D structure can create large changes in transcription. *eLife*, **10**, e64320.
15. Luppino,J.M., Park,D.S., Nguyen,S.C., Lan,Y., Xu,Z., Yunker,R. and Joyce,E.F. (2020) Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat. Genet.*, **52**, 840–848.
16. 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.

17. 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.
18. Fudenberg,G., Abdennur,N., Imakaev,M., Goloborodko,A. and Mirny,L.A. (2017) Emerging evidence of chromosome folding by loop extrusion. *Cold Spring Harb. Symp. Quant. Biol.*, **82**, 45–55.
19. Sanborn,A.L., Rao,S.S.P., Huang,S.-C., Durand,N.C., Huntley,M.H., Jewett,A.I., Bochkov,I.D., Chinnappan,D., Cutkosky,A., Li,J. *et al.* (2015) 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.
20. 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.
21. Hansen,A.S., Cattoglio,C., Darzacq,X. and Tjian,R. (2018) Recent evidence that tads and chromatin loops are dynamic structures. *Nucleus*, **9**, 20–32.
22. Rowley,M.J. and Corces,V.G. (2018) Organizational principles of 3D genome architecture. *Nat. Rev. Genet.*, **19**, 789.
23. Nuebler,J., Fudenberg,G., Imakaev,M., Abdennur,N. and Mirny,L.A. (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E6697–E6706.
24. 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.
25. Schwarzer,W., Abdennur,N., Goloborodko,A., Pekowska,A., Fudenberg,G., Loe-Mie,Y., Fonseca,N.A., Huber,W., Haering,C., Mirny,L. *et al.* (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, **551**, 51–56.
26. Kim,Y., Shi,Z., Zhang,H., Finkelstein,I.J. and Yu,H. (2020) Human cohesin compacts DNA by loop extrusion. *Science*, **366**, 1345–1349.
27. Bintu,B., Mateo,L.J., Su,J.-H., Sinnott-Armstrong,N.A., Parker,M., Kinrot,S., Yamaya,K., Boettiger,A.N. and Zhuang,X. (2018) Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, **362**, eaau1783.
28. Finn,E.H., Pegoraro,G., Brandão,H.B., Valton,A.-L., Oomen,M.E., Dekker,J., Mirny,L. and Misteli,T. (2019) Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell*, **176**, 1502–1515.
29. Mateo,L.J., Murphy,S.E., Hafner,A., Cinquini,I.S., Walker,C.A. and Boettiger,A.N. (2019) Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, **568**, 49.
30. Su,J.-H., Zheng,P., Kinrot,S.S., Bintu,B. and Zhuang,X. (2020) Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell*, **182**, 1641–1659.
31. Takei,Y., Yun,J., Zheng,S., Ollikainen,N., Pierson,N., White,J., Shah,S., Thomassie,J., Suo,S., Eng,C.-H.L. *et al.* (2021) Integrated spatial genomics reveals global architecture of single nuclei. *Nature*, **590**, 344–350.
32. Cardozo Gizzi,A.M., Cattoni,D.I., Fiche,J.-B., Espinola,S.M., Gurgo,J., Messina,O., Houbron,C., Ogiyama,Y., Papadopoulos,G.L., Cavalli,G. *et al.* (2019) Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol. Cell*, **74**, 212–222.
33. 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.
34. Mirny,L. and Dekker,J. (2021) Mechanisms of chromosome folding and nuclear organization: their interplay and open questions. *Cold Spring Harb. Perspect. Biol.*, **14**, a040147.
35. 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**, 496–501.
36. Mach,P., Kos,P.I., Zhan,Y., Cramard,J., Gaudin,S., Tünnermann,J., Marchi,E., Eglinger,J., Zuin,J., Kryzhanovska,M. *et al.* (2022) Cohesin and CTCF control the dynamics of chromosome folding. *Nat. Genet.*, **54**, 1907–1918.
37. Brandão,H.B., Gabriele,M. and Hansen,A.S. (2021) Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Curr. Opin. Cell Biol.*, **70**, 18–26.
38. Alexander,J.M., Guan,J., Li,B., Maliskova,L., Song,M., Shen,Y., Huang,B., Lomvardas,S. and Weiner,O.D. (2019) Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife*, **8**, e41769.
39. Fu,Y., Rocha,P.P., Luo,V.M., Raviram,R., Deng,Y., Mazzoni,E.O. and Skok,J.A. (2016) CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. *Nat. Commun.*, **7**, 11707.
40. Qin,P., Parlak,M., Kuscu,C., Bandaria,J., Mir,M., Szlachta,K., Singh,R., Darzacq,X., Yildiz,A. and Adli,M. (2017) Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nat. Commun.*, **8**, 14725.
41. Chen,B., Gilbert,L.A., Cimini,B.A., Schnitzbauer,J., Zhang,W., Li,G.-W., Park,J., Blackburn,E.H., Weissman,J.S., Qi,L.S. *et al.* (2013) Dynamic imaging of genomic loci in living Human cells by an optimized CRISPR/cas system. *Cell*, **155**, 1479–1491.
42. Germier,T., Audibert,S., Kocanova,S., Lane,D. and Bystricky,K. (2018) Real-time imaging of specific genomic loci in eukaryotic cells using the ANCHOR DNA labelling system. *Methods*, **142**, 16–23.
43. Thompson,A.P., Aktulga,H.M., Berger,R., Bolintineanu,D.S., Brown,W.M., Crozier,P.S., Veld,P.J., Kohlmeyer,A., Moore,S.G., Nguyen,T.D. *et al.* (2022) LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.*, **271**, 108171.
44. Woringer,M., Darzacq,X. and Izeddin,I. (2014) Geometry of the nucleus: A perspective on gene expression regulation. *Curr. Opin. Chem. Biol.*, **20**, 112–119.
45. Rippe,K. (2007) Dynamic organization of the cell nucleus. *Curr. Opin. Genet. Dev.*, **17**, 373–380.
46. Dekker,J. and Misteli,T. (2015) Long-range chromatin interactions. *Cold Spring Harb. Perspect. Biol.*, **7**, a019356.
47. Michaud-Agrawal,N., Denning,E.J., Woolf,T.B. and Beckstein,O. (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, **32**, 2319–2327.
48. Shao,S., Zhang,W., Hu,H., Xue,B., Qin,J., Sun,C., Sun,Y., Wei,W. and Sun,Y. (2016) Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Res.*, **44**, e86.
49. Arbona,J.-M., Herbert,S., Fabre,E. and Zimmer,C. (2017) Inferring the physical properties of yeast chromatin through bayesian analysis of whole nucleus simulations. *Genome Biol.*, **18**, 81.
50. Guo,Y., Al-Jibury,E., Garcia-Millan,R., Ntagiantas,K., King,J.W.D., Nash,A.J., Galjart,N., Lenhard,B., Rueckert,D., Fisher,A.G. *et al.* (2022) Chromatin jets define the properties of cohesin-driven in vivo loop extrusion. *Mol. Cell*, **82**, 3769–3780.
51. Banigan,E.J. and Mirny,L.A. (2020) The interplay between asymmetric and symmetric DNA loop extrusion. *eLife*, **9**, e63528.
52. Vian,L., Pękowska,A., Rao,S.S.P., Kieffer-Kwon,K.-R., Jung,S., Baranello,L., Huang,S.-C., Khattabi,L.E., Dose,M., Pruett,N. *et al.* (2018) The energetics and physiological impact of cohesin extrusion. *Cell*, **173**, 1165–1178.
53. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B (Methodological)*, **57**, 289–300.
54. Zhang,B., Zerubia,J. and Olivo-Marin,J.-C. (2007) Gaussian approximations of fluorescence microscope point-spread function models. *Appl. Opt.*, **46**, 1819–1829.
55. Virtanen,P., Gommers,R., Oliphant,T.E., Haberland,M., Reddy,T., Cournapeau,D., Burovski,E., Peterson,P., Weckesser,W., Bright,J. *et al.* (2020) SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**, 261–272.
56. 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.
57. Haering,C.H., Löwe,J., Hochwagen,A. and Nasmyth,K. (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell*, **9**, 773–788.
58. Srinivasan,M., Scheinost,J.C., Petela,N.J., Gligoris,T.G., Wissler,M., Ogushi,S., Collier,J.E., Voulgaris,M., Kurze,A., Chan,K.-L. *et al.* (2018) The cohesin ring uses its hinge to organize DNA using non-topological as well as topological mechanisms. *Cell*, **173**, 1508–1519.

59. 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.
60. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.
61. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
62. Kent,W.J., Zweig,A.S., Barber,G., Hinrichs,A.S. and Karolchik,D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, **26**, 2204–2207.
63. Amemiya,H.M., Kundaje,A. and Boyle,A.P. (2019) The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.*, **9**, 9354.
64. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-seq (MACS). *Genome Biol.*, **9**, R137.
65. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
66. Durand,N.C., Shamim,M.S., Machol,I., Rao,S.S.P., Huntley,M.H., Lander,E.S. and Aiden,E.L. (2016) Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Syst.*, **3**, 95–98.
67. Greenwald,W.W., Li,H., Smith,E.N., Benaglio,P., Nariai,N. and Frazer,K.A. (2017) Pgltools: a genomic arithmetic tool suite for manipulation of hi-C peak and other chromatin interaction data. *BMC Bioinf.*, **18**, 207.
68. Stansfield,J.C., Cresswell,K.G., Vladimirov,V.I. and Dozmorov,M.G. (2018) HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *BMC Bioinf.*, **19**, 279.
69. Natsume,T., Kiyomitsu,T., Saga,Y. and Kanemaki,M.T. (2016) Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep.*, **15**, 210–218.
70. Li,S., Prasanna,X., Salo,V.T., Vattulainen,I. and Ikonen,E. (2019) An efficient auxin-inducible degron system with low basal degradation in human cells. *Nat. Methods*, **16**, 866–869.
71. Banigan,E.J., van den Berg,A.A., Brandão,H.B., Marko,J.F. and Mirny,L.A. (2020) Chromosome organization by one-sided and two-sided loop extrusion. *eLife*, **9**, e53558.
72. Hsieh,T.-H.S., Cattoglio,C., Slobodyanyuk,E., Hansen,A.S., Rando,O.J., Tjian,R. and Darzacq,X. (2020) Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell*, **78**, 539–553.
73. Matthey-Doret,C., Baudry,L., Breuer,A., Montagne,R., Guiglielmoni,N., Scolari,V., Jean,E., Campeas,A., Chanut,P.H., Oriol,E. *et al.* (2020) Computer vision for pattern detection in chromosome contact maps. *Nat. Commun.*, **11**, 5795.
74. Thiecke,M.J., Wutz,G., Muhar,M., Tang,W., Bevan,S., Malysheva,V., Stocsits,R., Neumann,T., Zuber,J., Fraser,P. *et al.* (2020) Cohesin-dependent and -independent mechanisms mediate chromosomal contacts between promoters and enhancers. *Cell Rep.*, **32**, 107929.
75. 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.
76. Cattoglio,C., Pustova,I., Walther,N., Ho,J.J., Hantsche-Grininger,M., Inouye,C.J., Hossain,M.J., Dailey,G.M., Ellenberg,J., Darzacq,X. *et al.* (2019) Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *eLife*, **8**, e40164.
77. Herbert,S., Brion,A., Arbona,J.-M., Lelek,M., Veillet,A., Lelandais,B., Parmar,J., Fernández,F.G., Almayrac,E., Khalil,Y. *et al.* (2017) Chromatin stiffening underlies enhanced locus mobility after DNA damage in budding yeast. *EMBO J.*, **36**, 2595–2608.
78. Clow,P.A., Du,M., Jillette,N., Taghbalout,A., Zhu,J.J. and Cheng,A.W. (2022) CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci with one guide RNA per locus. *Nat. Commun.*, **13**, 1871.
79. Fudenberg,G. and Imakaev,M. (2017) FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat. Methods*, **14**, 673–678.
80. Finn,E.H. and Misteli,T. (2019) Molecular basis and biological function of variability in spatial genome organization. *Science*, **365**, eaaw9498.
81. Haarhuis,J.H.I., van der Weide,R.H., Blomen,V.A., Flach,K.D., Teunissen,H., Willems,L., Brummelkamp,T.R., Rowland,B.D. and de Wit,E. (2022) A mediator-cohesin axis controls heterochromatin domain formation. *Nat. Commun.*, **13**, 754.
82. 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.
83. Brandão,H.B., Ren,Z., Karaboja,X., Mirny,L.A. and Wang,X. (2021) DNA-loop-extruding SMC complexes can traverse one another in vivo. *Nat. Struct. Mol. Biol.*, **28**, 642–651.
84. Kim,E., Kerssemakers,J., Shaltiel,I.A., Haering,C.H. and Dekker,C. (2020) DNA-loop extruding condensin complexes can traverse one another. *Nature*, **579**, 438–442.
85. Krietenstein,N., Abraham,S., Venev,S.V., Abdennur,N., Gibcus,J., Hsieh,T.-H.S., Parsi,K.M., Yang,L., Maehr,R., Mirny,L.A. *et al.* (2020) Ultrastructural details of mammalian chromosome architecture. *Mol. Cell*, **78**, 554–565.
86. Miron,E., Oldenkamp,R., Brown,J.M., Pinto,D.M.S., Xu,C.S., Faria,A.R., Shaban,H.A., Rhodes,J.D.P., Innocent,C., Ornellas,S. de *et al.* (2020) Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *Sci. Adv.*, **6**, eaba8811.
87. 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.
88. Lelandais,B. and Sabaté,T. (2023) imodpasteur/LoopExtrusion: LoopExtrusion. https://doi.org/10.5281/ZENODO.7521634.
89. Sabaté,T., Lelandais,B., Bertrand,E. and Zimmer,C. (2023) Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging. https://doi.org/10.5281/ZENODO.7525055.

Supplementary Figures

Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging

Thomas Sabaté[1,2,3,*] , Benoît Lelandais[1,4], Edouard Bertrand[2,†], Christophe Zimmer[1,*,†]

1. Institut Pasteur, Université Paris Cité, Imaging and Modeling Unit, F-75015 Paris, France

2. IGH, University of Montpellier, CNRS UMR 9002, Montpellier, France

3. Sorbonne Université, Collège Doctoral, F-75005 Paris, France

4. Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France

† These authors contributed equally to this work.

* To whom correspondence should be addressed: Thomas Sabaté and Christophe Zimmer

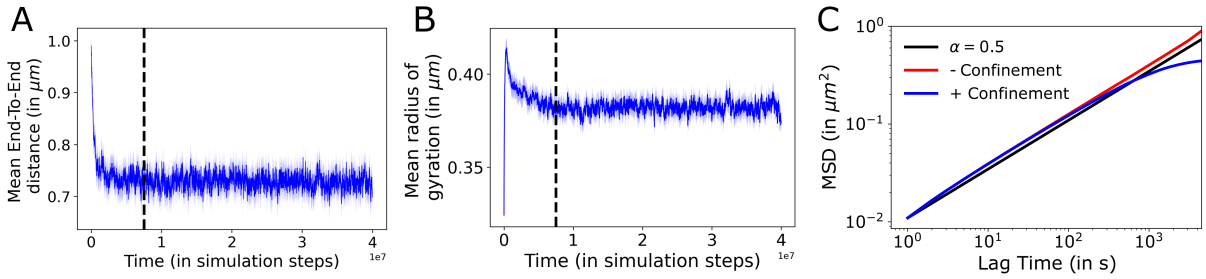


Figure S1: Equilibration and Rouse dynamics of simulated polymers

A-B: Mean end-to-end distance (**A**) and mean radius of gyration (**B**) of simulated polymers as function of simulation time steps (the means are computed over n=1,000 independent simulations). Black dotted lines show the beginning of conformation snapshots recorded for subsequent analyses. The blue shaded area shows the 95% confidence interval. **C:** Mean-Squared Displacements (MSD) as function of lag time for bead #300 of a simulated polymer with 600 beads. Blue and red curves correspond to polymers with confinement (in a sphere of radius 18 beads) or without confinement, respectively. The MSD is proportional to the square root of lag time at small time scales, consistent with the Rouse regime expected for an ideal polymer (black line) (1). The MSD reaches a plateau at long time scales due to polymer confinement.

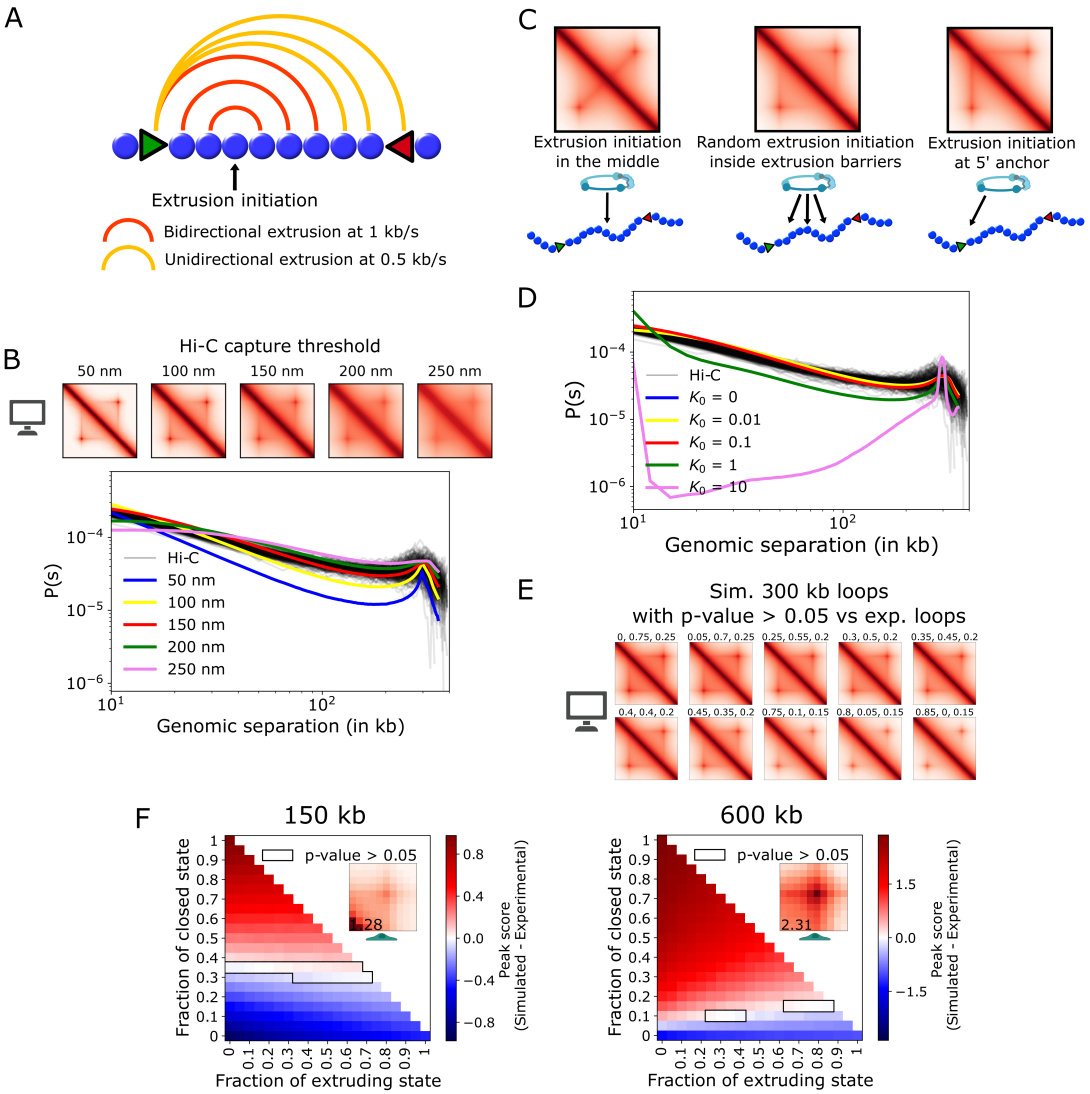


Figure S2: Comparison of polymer simulations to experimental Hi-C data

A: Sketch of the loop extrusion model showing bidirectional extrusion at 1 kb/s until one of the two anchors is reached, followed by unidirectional extrusion of the other side of the polymer at 0.5 kb/s until the second anchor is reached. The arcs show pairs of beads kept in close proximity during the extrusion process. **B:** Simulated contact maps for different capture radii thresholds (shown with the same display range) and the corresponding *P(s)* curves (average contact frequency as function of genomic separation). Black curves show the experimental *P(s)* for loops of size 300 +/- 30 kb. **C:** Simulated contact maps for various conditions of extrusion initiation. **D:** Simulated *P(s)* curves for polymers with different stiffness parameters K_0. $\mathrm{K}_0 = 0.1$ was used for all simulations analyzed elsewhere. Black curves as in **B**. Simulated *P(s)* curves in **B** and **D** were shifted vertically to facilitate comparison. In **B**, **C** and **D**, the simulated contact maps were generated assuming the following loop state fractions: 40% open, 40% extruding and 20% closed states. **E:** Simulated contact maps with peak scores consistent with the experimental Hi-C data (statistical test with p-value>0.05) for loops of 300 kb (+/- 30 kb). Different fractions of loop states can recapitulate experimental peak scores. The fraction of each loop state is indicated above each map in the following order: open, extruding, closed. **F:** Same as **Fig. 2C**, but for loops of 150 kb +/- 15 kb (n=179, left) and 600 kb +/- 60 kb (n=91, right). Contact maps in **B**, **C**, **E** are shown at 3 kb resolution, the two aggregated contact maps in **F** are shown at 15 kb resolution, using a different display range.

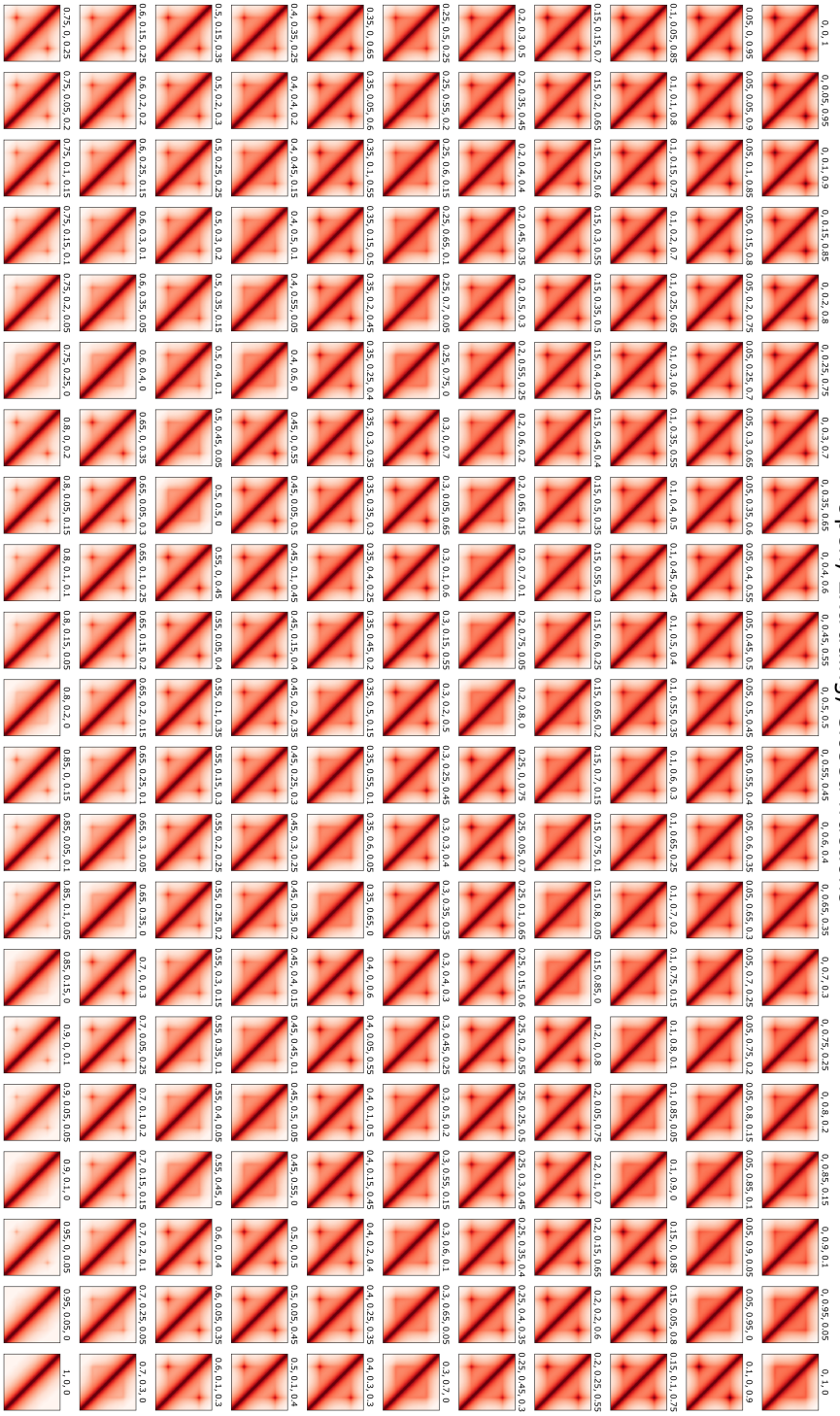


Figure S3: Simulated contact maps for all combinations of loop state fractions.
Simulated contact maps based on simulations with different combinations of the three loop state fractions, shown at 3 kb resolution. The fractions of each loop state are indicated above each map in the following order: open, extruding, closed. Top right: closed states only. Bottom right: extruding states only. Bottom left: open states only.

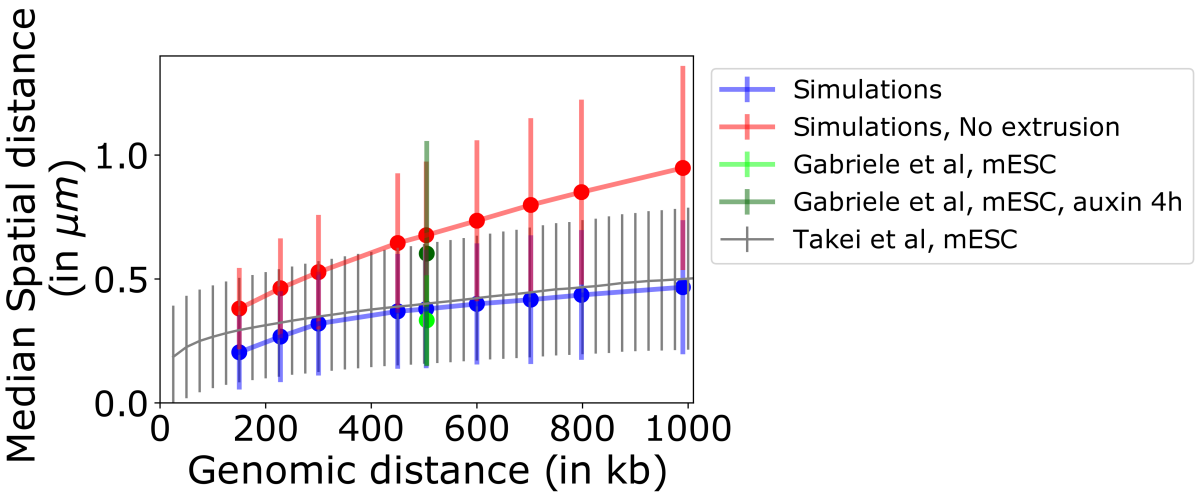


Figure S4: Comparison of simulations to imaging data.

Median spatial distance between pairs of loci as function of genomic distance for simulated polymers with (blue) or without (red) loop extrusion, and measured in experimental imaging data (grey and green). The grey curve shows DNA-FISH measurements in mouse embryonic stem cells (mESC) by Takei et al (2). This study did not specifically target loop anchors. The green dots show the distance between two loop anchors measured in live mESC cells by Gabriele et al (3) (clone 36) without auxin (light green) or with a 4h auxin treatment leading to RAD21 depletion (dark green). For the simulations, we sampled anchor-anchor distances using mean loop state fractions consistent with experimental Hi-C data and assumed random localization errors consistent with Takei et al (2) ($\sigma_{x,y} = 41$ nm, $\sigma_z = 65$ nm). Error bars show standard deviations.

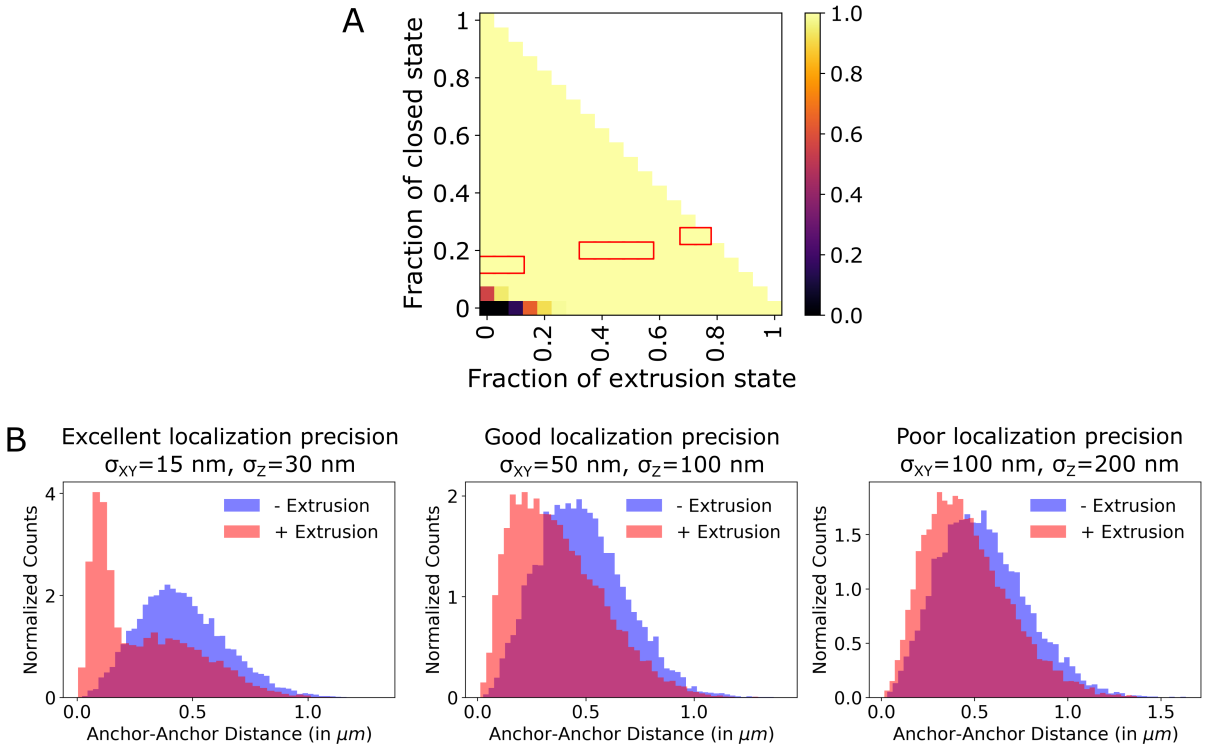


Figure S5: Detectability of loop extrusion for large samples and realistic experimental parameters. **A:** Same as the heat map in **Fig. 3E**, but for N=1,000 (instead of N=100) measurements of anchor-anchor distances. **B:** Distribution of anchor-anchor distances in simulations with (red) and without (blue) loop extrusion for excellent, good and poor localization precisions (as indicated), assuming the following realistic parameters: reporter-anchor separation: 3 kb on each side (6 kb total); extrusion speed: 1 kb/s; loop size: 300 kb; sample size: N=10,000 distances; loop state fractions: 40% open, 40% extruding and 20% closed. Lateral and axial localization precisions are indicated above each plot. A peak at low distances due to the closed state is visible only for excellent localization precision.

Parameter	**Our simulations**	**Gabriele et al, Science, 2022**
Loop length	504 kb	505 kb (Fig. 1C)
Reporter - Anchor separation	2 x 6 kb	2 x 5 kb (Fig. 1C)
Extrusion speed	1 kb/s	unknown
Loop state fractions	2% closed, 92% extruding, 6% open	2% closed, 92% extruding, 6% open (Fig. 4D)
Localization precision	With extrusion: σ_{xy}=47 nm σ_z=120 nm Without extrusion: σ_{xy}=44 nm σ_z=129 nm	WT (C36): σ_x=48 nm σ_y=47 nm σ_z=120 nm Upon cohesin degradation: σ_x=45 nm σ_y=44 nm σ_z=129 nm (Table S3)
Sample size	Variable, from N=10 to N=45,000	N>45000

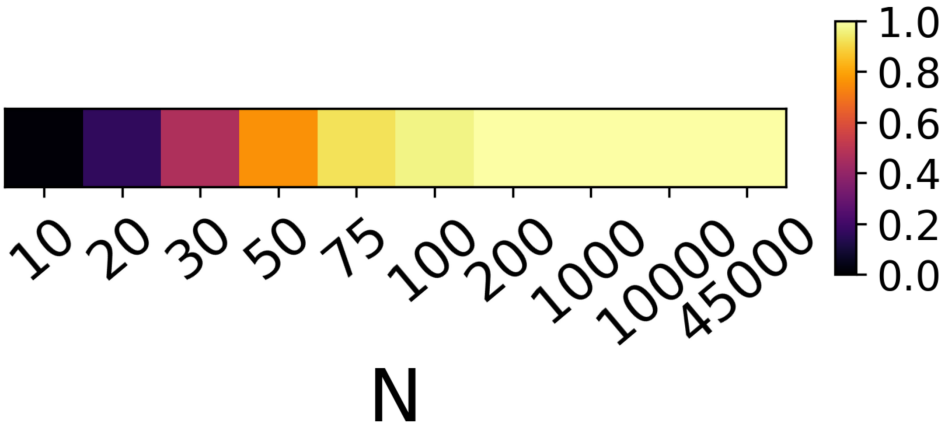


Figure S6: Detectability of loop extrusion using experimental parameters from Gabriele et al.
The color bar (bottom) shows loop extrusion detectability (defined as in **Fig. 3**) for different sample sizes (number of distance measurements) N, based on simulations using parameters similar to those reported in the experimental study by Gabriele et al. (3), as shown in the table (top).

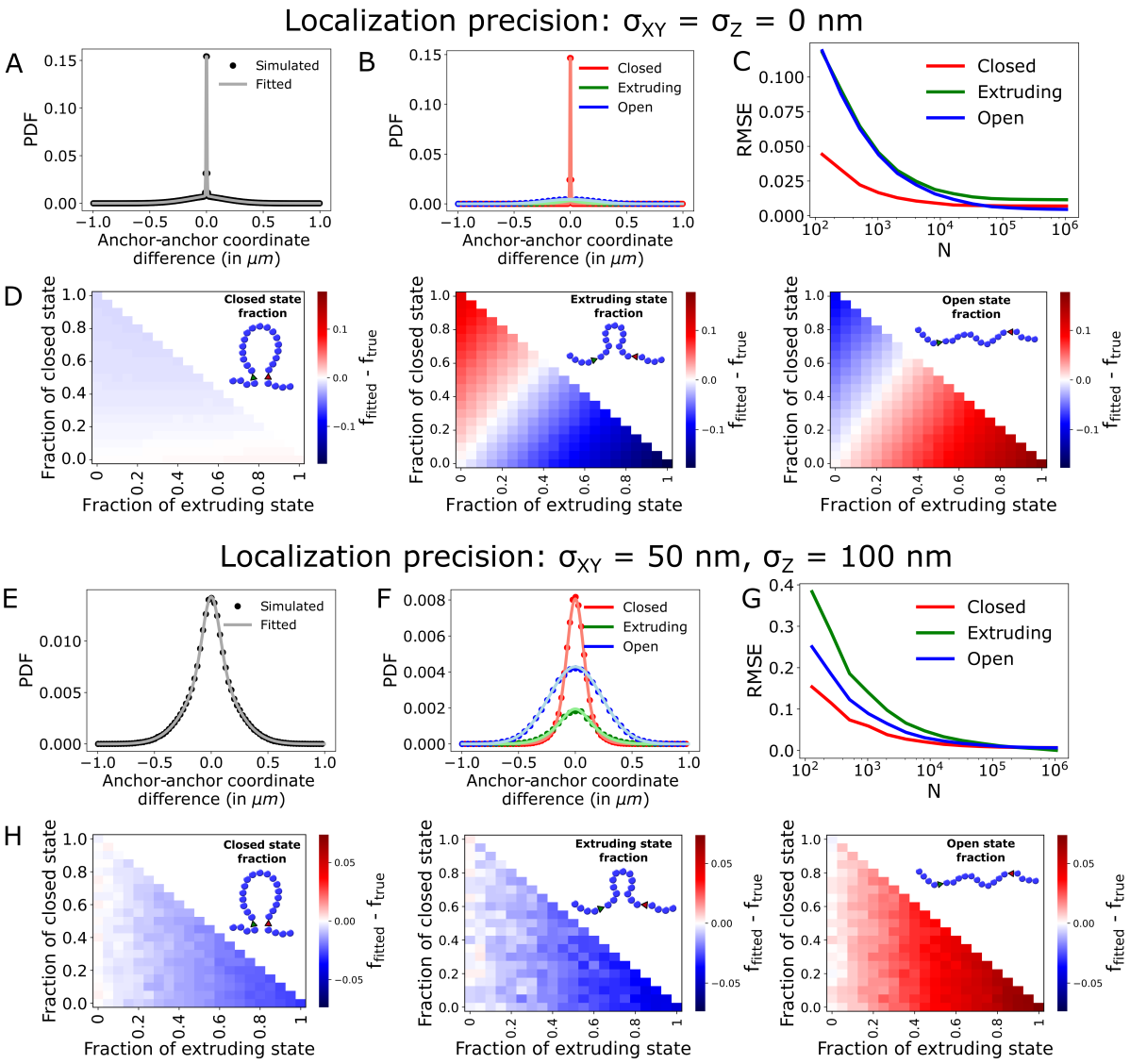


Figure S7: Quantifying fractions of loop states from static images for different localization precisions. **A-D**: Same as **Fig. 4** but without localization errors ($\sigma_{x,y} = \sigma_z/2 = 0$ nm). **E-H**: Same as Fig. 4, but with larger localization errors ($\sigma_{x,y} = \sigma_z/2 = 50$ nm).

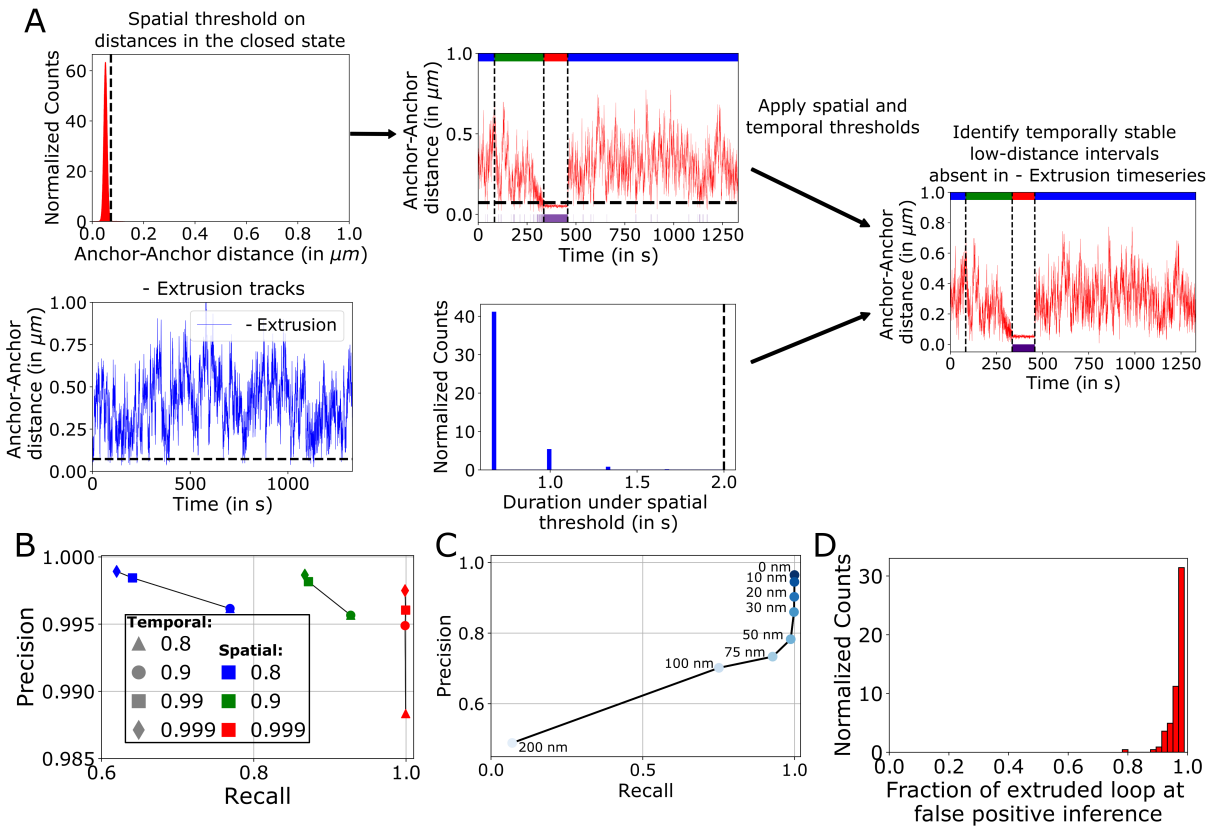


Figure S8: Segmentation of closed loop states from anchor-anchor distance time series.
A: Pipeline to segment closed loop states. First, a spatial threshold was defined as a quantile of anchor-anchor distances in the closed state. Second, a temporal threshold was defined as the number of consecutive frames with distances below the spatial threshold in time series without extrusion. The two thresholds were then used together to define closed state segments as stable low-distance time intervals absent from extrusion-free time series. As in **Fig. 5A**, colored bars on top indicate the true loop state (blue: open, green: extruding, red: closed), while the indigo bar at the bottom indicates a segmented closed state. **B:** Precision and recall of segmented closed state timepoints for varying spatial and temporal thresholds. The spatial quantiles used to define the temporal threshold are indicated in different colors, while the temporal thresholds are indicated with different shapes. **C:** Precision and recall of segmented closed state timepoints using 0.999 quantiles as spatial and temporal thresholds for different localization precisions and with a total reporter-anchor separation of 6 kb, mimicking realistic experimental conditions. Assumed localization errors $\sigma_{x,y} = \sigma_z/2$ are indicated for each data point. **D:** Fraction of the genomic distance between the anchors that was extruded at the start of falsely detected closed states (0 means that extrusion did not start, 1 means that the two anchors were brought into contact). False positives are defined as either: (i) segmented closed state intervals that did not overlap any true closed interval, or (ii) if two or more segmented intervals overlapped a single true interval, the longest segmented interval was defined as the true positive and the smaller interval(s) as false positive(s). Only false positive closed states detected during extrusion (which accounted for 95% of false positives) were plotted.

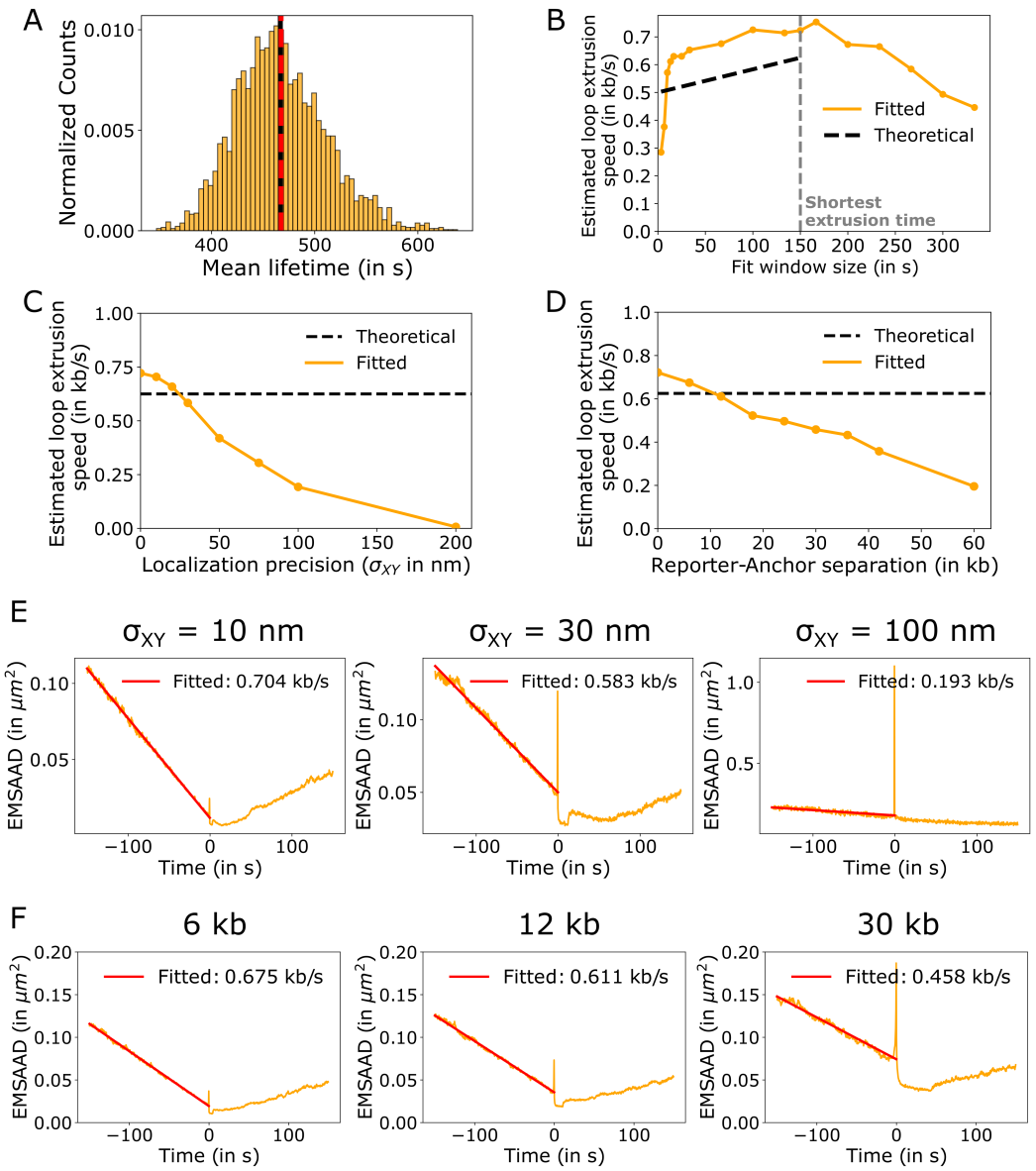


Figure S9: Estimation of closed state lifetimes by bootstrapping and estimation of extrusion speed. **A:** Histogram shows the distribution of the estimated mean closed state lifetimes based on 4,000 bootstrapped simulated samples of 1,000 time series each. The red dotted and black continuous lines show the estimated and theoretical mean lifetimes, respectively (the two lines overlap). **B:** The orange curve shows the estimated mean effective loop extrusion speed as function of the time interval before the start time of the segmented closed states. The black dashed line shows the theoretical extrusion speed. The grey vertical line shows the minimal duration of extrusion (corresponding to extrusion initiation at the mid-point between the two anchors). To the left of this line, all time series are in the extruding state; to the right, some time series are in the open state and others in the extruding state. **C-D:** Orange curves show the mean estimated loop extrusion speed from segmented closed states as function of localization precision (**C**) or reporter-anchor separation (**D**). The black dotted line indicates the theoretical average loop extrusion speed. **E-F**: Ensemble mean squared anchor-anchor distance (EMSAAD) obtained by averaging and aligning time series (orange curves), are shown together with a linear fit (red line) for different localization precisions (**E**) or total reporter-anchor separations (**F**).

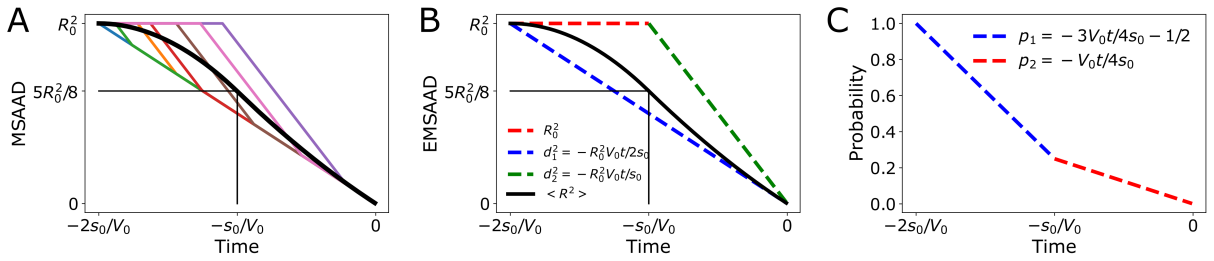


Figure S10: Theoretical model of mean squared anchor-anchor distances with loop extrusion.
This Figure describes theoretical models for the mean squared anchor-anchor distance (MSAAD) or the ensemble MSAAD (EMSAAD) as function of time based on an analytical polymer model with loop extrusion (see Materials and Methods). No molecular dynamics simulations were used here. **A:** The seven colored curves show the theoretical MSAAD as function of time, with each curve corresponding to a different site of extrusion initiation (due to random landing of the extrusion complex between the two anchors). Each curve consists of three linear parts, because the instantaneous extrusion speed (in base pairs per second) varies from $V = 0$ (before extrusion begins), to $V = V_0$ (bidirectional extrusion, before reaching the first anchor), then to $V = V_0/2$ (unidirectional extrusion, after one of the two anchors has been reached). The genomic distance between the two anchors is s_0. The black curve shows the EMSAAD assuming that extrusion is initiated with uniform random probability between the two anchors. The EMSAAD was obtained by aligning 10^4 individual time traces (such as the seven colored curves) at the beginning of the closed state ($t = 0$) and averaging. **B:** Theoretical EMSAAD as function of time for averaged and aligned time traces as predicted by Eq. 4 (black). Note that this theoretical model perfectly matches the black curve in **A**. The colored dashed curves indicate lower and upper bounds corresponding to a constant extrusion speed of $V = 0$ (red), $V = V_0/2$ (blue) or $V = V_0$ (green). **C:** Dashed lines indicate probabilities as function of time that the MSAAD equals R_0^2 (p_1, blue), or equals $d_1^2 = -R_0^2 V_0 t/(2s_0)$ ($1 - p_1$, for $t \leq -s_0/V_0$ and $1 - p_2$, for $t \geq -s_0/V_0$) or equals $d_2^2 = -R_0^2 V_0 t/s_0$ (p_2, red).

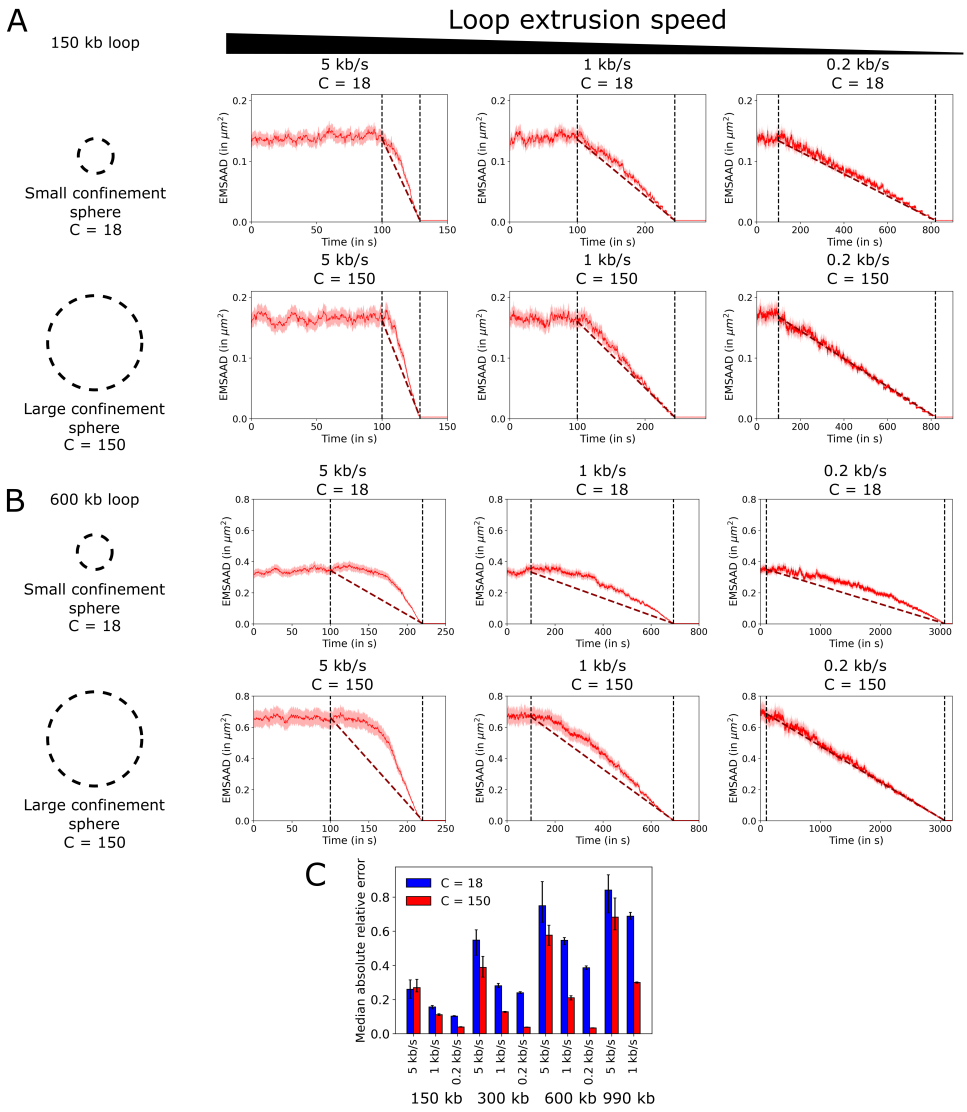


Figure S11: Deviation of extruding polymer simulations from theoretical model at equilibrium.
A-B: Red curves show the ensemble mean squared anchor-anchor distance (EMSAAD) as function of time, averaged over 400 synchronized trajectories for different speeds of extrusion (0.2 kb/s, 1 kb/s and 5 kb/s), different radii of confining spheres (C=18 or C=150 bead diameters) and for loops of 150 kb (**A**) or 600 kb (**B**). For these simulations, extrusion was assumed to initiate at the mid-point between the two anchors, therefore proceeding bidirectionally until both anchors are reached simultaneously. The first and second vertical black dotted lines show the start and end of extrusion respectively. The dark red dashed line shows the theoretical (linear) EMSAAD for equilibrated polymers assuming a constant extrusion velocity (Eq. 3 and **Fig. S10B**, green dashed line). The light red area shows the 95% confidence interval. Simulations matched the theoretical model for low extrusion speeds and weak confinement (e.g. $V_0 = 0.2$ kb/s and C=150 bead diameters), whereas high extrusion speed and/or confinement (e.g. $V_0 = 5$ kb/s and C=18 bead diameters) led to deviations from the linear model. **C:** Median absolute relative error between theoretical and simulated EMSAAD. Blue and red bars indicate the radius of the confining sphere (C=18 and C=150 bead diameters, respectively). The error was defined as $| y_{\text{theo}} - y_{\text{sim}} | / y_{\text{theo}}$ for every timepoint in the extruding state, where the subscripts 'theo' and 'sim' refer to theoretical and simulated values, respectively. Error bars show the 95% confidence interval.

REFERENCES

1. Rosa,A. and Zimmer,C. (2014) Chapter Nine - Computational Models of Large-Scale Genome Architecture. In Hancock,R., Jeon,K.W. (eds), *International Review of Cell and Molecular Biology*, New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals. Academic Press, Vol. 307, pp. 275–349.

2. Takei,Y., Yun,J., Zheng,S., Ollikainen,N., Pierson,N., White,J., Shah,S., Thomassie,J., Suo,S., Eng,C.-H.L., *et al.* (2021) Integrated spatial genomics reveals global architecture of single nuclei. *Nature*, **590**, 344–350.

3. 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**, 496–501.

1.2 Polymer simulations define the minimal requirements for quantifying chromatin loop extrusion

1.2.1 Summary of polymer simulations results

We modelled loop extrusion on a generic fragment of chromatin, modelled as a polymer undergoing Langevin dynamics. By comparing experimental and simulated contact maps, I could estimate the fraction of each loop state that was consistent with our simulations. This estimation was necessary to analyze anchor-anchor distance distributions in realistic conditions since they are heavily impacted by varying fractions of loop states (**Figure 2** of Sabaté *et al*, NAR, 2023). I then defined the minimal experimental requirements to detect chromatin loop extrusion by imaging loop anchors. It was found that loop extrusion can generally be detected, using realistic levels of noise and sampling size. For instance, mixing all sources of noise together, loop extrusion could still be detected with N=200 measurements of anchor-anchor distances and with a poor localization precision of $\sigma_{XY} = \frac{\sigma_Z}{2} = 100$ nm (**Figure 3** of Sabaté *et al*, NAR, 2023).

Next, we developed an analytical method to estimate the fraction of loop states from static imaging (*e.g.* DNA FISH) of loop anchors using cells with and without cohesin. The decomposition of the total anchor-anchor distance distribution into three different distributions corresponding to the closed, extruding and open states is based on two main assumptions. First, in the extruding state, the anchors behave as if part of a shorter polymer from which the extruded loop is absent. Second, the polymer has time to equilibrate at each extrusion step. Based on this model, we could determine with an overall good performance (absolute bias lower than 13%) the fraction of each state (**Figure 4** of Sabaté *et al*, NAR, 2023). Being able to estimate these parameters from static imaging alone is powerful since it requires much less experimental work to measure anchor-anchor distance in fixed rather than in living cells. In addition, hundreds or thousands of loop anchors could be targeted by using sequential FISH[134,159,286] or high-throughput DNA FISH[132] to estimate these parameters genome-wide and define how loop structures vary in different chromatin environments.

Finally, we segmented closed states in dynamic tracking of loop anchors to estimate the fraction of closed states and their lifetime (**Figure 5** of Sabaté *et al*, NAR, 2023). As opposed to published methods of closed state segmentation[163,164], our method does not rely on any assumptions (except that the anchor-anchor distance is higher in cohesin-depleted cells)

or models. Our closed state segmentation had good performance for low to moderate (73% precision and 93% recall for $\sigma_{X,Y} = \frac{\sigma_Z}{2} = 75$ nm and 6 kb of reporter-anchor genomic separation) levels of noise. Based on this closed state segmentation and the above-mentioned assumptions, we could accurately fit the effective loop extrusion speed from the temporal tracks (**Figure 5** of Sabaté *et al*, NAR, 2023). Applied to experimental data, this method would provide a completely novel information on the *in vivo* loop extrusion speed.

Apart from validating analytical methods to quantify loop extrusion, polymer simulations allowed us to define crucial controls for this quantification. Indeed, two controls were necessary to make these estimations:

- A closed state control used to approximate the closed state.

- An open state control where all loops are eliminated. This can be obtained by depleting a cohesin subunit such as RAD21[31] or the cohesin loader NIPBL[156].

1.2.2 Limitations and possible improvements

This polymer simulation analysis also comes with some limitations (mentioned in the Discussion of Sabaté *et al*, NAR, 2023). I highlight some of them below:

Minimal experimental requirements to detect loop extrusion.

The experimental conditions defined to detect loop extrusion are likely to be optimistic. Indeed, in our simulations the links between beads were solely created from the loop extrusion process. However, apart from loop extrusion, several other processes might create long-range interactions in the nucleus (compartmentalization, transcription, phase separation[75,287]), which would make loop extrusion more difficult to detect. Moreover, the nature or strength of these links might be modified in RAD21-depleted cells (*e.g.* strengthened compartmentalization in RAD21-depleted cells[31,44,61,156]) as compared to WT cells, which would complicate the detection of loop extrusion-dependent changes in anchor-anchor distance.

Analytical models and their assumptions.

The analytical models assume two main hypotheses. First, in the extruding state, the anchors behave as if part of a shorter polymer from which the extruded loop is absent.

Second, the polymer has time to equilibrate at each extrusion step. In our simulations, loop extrusion did not proceed at equilibrium for rates higher than 0.2 kb/s and both assumptions are likely to be violated in the cells. Deriving out-of-equilibrium models would require extensive work, while the current model still accurately estimated the fitted parameters despite at least one assumption being violated in the simulations.

Simplicity of loop extrusion modeling.

In our simulations, we modelled loop extrusion with a single cohesin molecule extruding the loop. Cohesin exhibited two different speeds, did not pause and did not unbind before reaching the anchors. A larger diversity of scenarios could be found in the cells and would complicate the analysis.

We defined the minimal experimental design requirements to detect loop extrusion and validated analysis methods to quantify loop extrusion in living cells. Despite the mentioned limitations, these analysis methods should be applicable to experimental data, using appropriate controls.

2. Tracking loop anchors in living cells

2.1 Visualizing specific chromatin loci with repeat arrays

Among the available techniques to visualize specific chromatin loci in living cells, I chose to use repeat arrays (such as lacO, tetO or CuO) bound by fluorescent proteins (**Table 1** and **Figure 14**). These repeat arrays represent an established and efficient method to visualize chromatin[259]. The fluorescent signal and repeat array length can be easily tuned depending on the number of repeats, which is not possible with the ANCHOR system. Two orthologous repeats were used to enable two-color imaging of the anchors. Among the three main repeat types available, I chose to use tetO and CuO because lacO repeats were reported to induce chromosomal instability in mESCs[288].

At the time that I designed the strategy to visualize chromatin loci, dCas9 was still in development and seemed difficult to use for targeting non repetitive loci[265–267,289]. Although dCas9 imaging of specific chromatin loci is extremely promising due to its theoretical ease of use and the numerous method improvements, it is still challenging to image non-repetitive chromatin loci[269].

Despite their advantages, repeat arrays are repetitive sequences that may recombine during DNA replication and therefore fluorescent signal might be lost upon cell divisions. Furthermore, the impact on chromatin motion and state of inserting long sequences bound by transcription factors, as well as on interference with loop extrusion remains poorly characterized. Finally, the significant length of the arrays required to provide sufficient signal makes their CRISPR-mediated insertion in the genome difficult. I describe below how I tackled some of these challenges.

2.2 Difficulty of inserting long repeat array in specific genomic regions by CRISPR-Cas9

2.2.1. Low expected rate of repeat array insertion by CRISPR-Cas9

The visualization of a pair of chromatin loop anchors was made possible by inserting two orthogonal repeat arrays (TetOx96 and CuOx150) in specific regions of the genome using CRISPR-Cas9.

The insertion of our repeat arrays is complicated by several reasons. First, because the intensity of the fluorescent spot depends on the number of repeats, long segments of DNA (6 and 8 kb) must be inserted. However, the length of the donor DNA was found to anti-correlate with the efficiency of CRISPR-mediated genome insertion[274,290]. Second, homology arms of at least 500 bp are necessary to reach good insertion efficiency[274,278,290]. In addition to the long sequences of our repeat array, this hinders the use of PCR DNA donors, which are known to increase HDR efficiency during CRISPR genome editing as compared to plasmid donors[278]. Third, I label endogenous loops and need to be as close as possible to the CTCF site defining the loop anchor (based on the polymer simulation analysis of reporter-anchor genomic separation, **Figure 3C** of Sabaté *et al*, NAR, 2023), limiting the choice of useable sgRNAs. Fourth, an antibiotic-resistance gene used for selection is expressed by its own promoter. In contrast with gene labeling where the resistance marker is under the expression of the endogenous targeted gene, any incorrect integration of our cassette within the genome will drive the resistance gene expression. Therefore, a higher number of incorrect clones will grow despite the antibiotic selection, as compared to traditional gene editing.

Thus, the length of our repeat array is likely to decrease its CRISPR integration rate. However, by using long homology arms and optimizing sgRNA design, I expect to reach a sufficiently high HDR efficiency to label several loops in the genome.

CRISPR is known to be error-prone and to display high heterogeneity in integration outcomes. Insertions can be made at non-desired locations (off-target events) or with truncated, extended or duplicated integrations[278]. These undesired events can be filtered out by (1) imaging: off-target events yield fluorescent spots several micrometers away from each other, which is not expected from loci separated by a few hundreds of kb[134,286] (2) PCR genotyping allows to filter out incorrect on- and off-target integrations.

2.2.2 Degenerated repeats reduce the risk of repeat recombination

Repeat arrays typically comprise 100-200 repetitions of short sequences (about 20 bp). The repetition of the same sequence can trigger recombination and thus reduce the length of the array, which would reduce the signal of the visualized spot. To reduce this risk, both in bacteria during molecular cloning and in cells during CRISPR-induced HDR or DNA replication, I used degenerated sequences between each repeat. Repeats within a cluster of repeats were separated by short random sequences, while clusters of repeats were separated

by longer random sequences. All steps of the molecular cloning process involving bacteria amplification were carried out at 30°C to ensure integrity of the whole repeat array.

2.2.3 Sequential clonal selection is a tedious process

Inserting exogenous DNA in the genome requires to select clones that integrated the exogenous DNA and to carefully check if insertions are correct. The final cell population must emerge from a single clone to ensure the reproducibility of the results. Because human cells have doubling times of about 20h, the process of clone selection, clone genotyping and clone expansion takes at least 50 days for each genome insertion and thus requires an extensive amount of cell handling.

We decided to insert our two orthogonal repeats sequentially rather than simultaneously. Although the simultaneous insertion of repeat arrays seems faster because it requires only one round of clonal expansion (about 50 days) instead of two for sequential insertions, we cannot ensure that a clone with a double correct insertion will be identified among all clones. Moreover, recombination can occur between co-transfected plasmids of the repeat arrays. This could lead to incorrect insertions in the genome and further decrease the fraction of selected clones that can be used for imaging. For these reasons, I decided to insert our repeat arrays one after the other. Due to the lower efficiency of repeat insertion (see above-mentioned reasons), it was difficult to obtain homozygote clones. Therefore, after a first heterozygote insertion on one anchor, half of the correct insertions of the other anchor can occur on a different allele. This decreased by 50% the fraction of clones with correct insertions that could be used for imaging during the second CRISPR-mediated insertion.

Our goal was to track and quantify multiple chromatin loops and TADs throughout the genome by inserting long repeat arrays in specific regions of the genome using CRISPR-Cas9. For this aim, I needed to generate multiple repair plasmids containing the repeat array surrounded by the two homology arms for each specific genomic locus. The lower expected CRISPR-mediated insertion efficiency, the multiplicity of the repair cassettes to be generated and the tedious process of sequential clone selection and genotyping prompted us to carefully design an efficient, versatile, and robust molecular cloning strategy to generate multiple repair cassettes.

2.3 An efficient molecular cloning strategy to generate numerous CRISPR repair cassettes

Chapter 13

Versatile CRISPR-Based Method for Site-Specific Insertion of Repeat Arrays to Visualize Chromatin Loci in Living Cells

Thomas Sabate, Christophe Zimmer, and Edouard Bertrand

Abstract

Hi-C and related sequencing-based techniques have brought a detailed understanding of the 3D genome architecture and the discovery of novel structures such as topologically associating domains (TADs) and chromatin loops, which emerge from cohesin-mediated DNA extrusion. However, these techniques require cell fixation, which precludes assessment of chromatin structure dynamics, and are generally restricted to population averages, thus masking cell-to-cell heterogeneity. By contrast, live-cell imaging allows to characterize and quantify the temporal dynamics of chromatin, potentially including TADs and loops in single cells. Specific chromatin loci can be visualized at high temporal and spatial resolution by inserting a repeat array from bacterial operator sequences bound by fluorescent tags. Using two different types of repeats allows to tag both anchors of a loop in different colors, thus enabling to track them separately even when they are in close vicinity. Here, we describe a versatile cloning method for generating many repeat array repair cassettes in parallel and inserting them by CRISPR-Cas9 into the human genome. This method should be instrumental to studying chromatin loop dynamics in single human cells.

Key words Live-cell imaging, Repeat array, Chromatin 3D architecture, TADs, Loop extrusion

1 Introduction

The combination of Hi-C [1] or related genomic techniques and imaging methods such as DNA FISH [2–4] led to a growing understanding of the 3D architecture of the genome and to the discovery of chromatin structures including topologically associating domains (TADs) and chromatin loops [5, 6]. It has been shown that these structures emerge, at least partly, from a DNA loop extrusion process driven by the cohesin complex [7, 8]. However, most studies have been restricted to fixed cells and/or to cell population averaged snapshots of the genome architecture. The temporal dynamics of loops and TADs in single cells, although of major importance to understanding their role in gene regulation and chromosome organization, has been insufficiently characterized [9–11]. An important requirement to study the dynamics of

Tom Sexton (ed.), *Spatial Genome Organization: Methods and Protocols*,
Methods in Molecular Biology, vol. 2532, https://doi.org/10.1007/978-1-0716-2497-5_13,

chromatin regions by microscopy is to visualize specific genomic loci with live-cell compatible fluorescent labels. This can be achieved in various ways: using sgRNAs and catalytically dead Cas9 (dCas9) [12], using fluorescent reporter operator systems (FROS), such as lacO, TetO, or CuO [13, 14] arrays, or using the parS/ParB system [15]. Here, we describe a general cloning method to generate repair cassettes of CuO repeat arrays and to insert them in the genome of human cells by CRISPR-Cas9. This method can be extended to all types of repeat arrays for multicolor imaging.

Inserting the repeat array at a nontranscribed region of the genome will lead to more off-targets than tagging a coding sequence, because for nontranscribed regions the expression of the antibiotic resistance gene is under its own promoter while for a coding region it can be transcribed from the endogenous promoter via the use of an Internal Ribosome Entry Site (IRES) or 2A self-cleaving peptides. Therefore, we use a Cas9 D10A nickase that limits the off-target rate compared to a WT Cas9 [16]. We also use an improved scaffold of guide RNA that vastly increases their efficiency [12].

CRISPR-Cas9–mediated insertion of repeat arrays in the genome of human cells requires addition of homology arms on both sides of the repeat array. This can be achieved by PCR amplification of the long repeat array with primers already containing the homology arms [17]. However, this technique can only produce short homology arms (<200 bp), which reduces the efficiency of homology-directed repair insertion by Cas9, depends on the PCR yield and limits the choice of insertion sites. Therefore, we instead clone the homology arms into the repeat array plasmid. This procedure is longer than with PCR amplified repair cassettes but allows generation of longer homology arms (>800 bp). Hence, it improves the integration rate compared to PCR-amplified products and increases the choice of insertion sites. The combination of the Cas9 nickase, optimized guide scaffold, and long homology arms allows for large operator repeats to be integrated relatively easily and with a good efficiency (10–65% of antibiotic-resistant clones contain the repair cassette inserted at the right location depending on the targeted locus).

The cloning method consists of PCR amplification of the homology arms by genomic PCR, which are then integrated by Gibson assembly at each side of the repeat array. This method allows to generate quickly and in parallel multiple repair cassettes for specific tagging of multiple chromatin loci, increasing the throughput of such experiments. Using two different types of repeats enables two-color imaging, which allows localization and tracking of two loci with subpixel and subdiffraction precision, even when the distance between them falls below the optical resolution.

Thus, it is possible to analyze and quantify the movements and distances between anchors of TADs or chromatin loops in living cells. We propose that this protocol is a useful tool to study the dynamics of chromatin structures.

2 Materials

1. A U6-based expression plasmid containing an improved scaffold for guide RNAs (12; available upon request). The plasmid bears two mutations compared to the original sequence. First, a run of four Ts after the guide sequence is modified to prevent Pol-III termination. Second, a stem-loop in the guide scaffold is stabilized to favor its formation at 37 °C (Fig. 1a). The guides are designed such that they do not cleave the repair cassette (Fig. 1b).
2. 10 U/μL BpiI enzyme (provided with 10× buffer G, ThermoFisher).
3. Nuclease-free water.
4. Gel extraction and PCR purification kits (e.g., NucleoSpin Gel and PCR Clean-up; Macherey-Nagel).
5. NanoDrop spectrophotometer or equivalent.
6. Separately synthesized forward and reverse oligos containing the sgRNA sequences resuspended at 1 μg/μL in TE buffer (10 mM Tris–HCl pH 8, 1 mM EDTA). *See* Subheading 3.1 for design strategy.
7. 10 mM ATP.
8. 100 mM DTT.
9. 400 U/μL T4 DNA ligase (NEB).
10. Electrocompetent, recombinant-deficient *E. coli* (e.g., XL1-Blue).
11. Bacterial electroporation system.
12. SOB medium.
13. LB medium.
14. 100 μg/mL ampicillin LB agar plates.
15. 100 μg/mL ampicillin.
16. Plasmid miniprep kit (e.g., QIAprep miniprep kit; Qiagen).
17. 20 U/μL BbsI-HF, provided with 10× CutSmart buffer (NEB).
18. 20 U/μL ScaI-HF (NEB).
19. Plasmid midiprep kit (e.g., Nucleobond Xtra Midi kit; Macherey-Nagel).

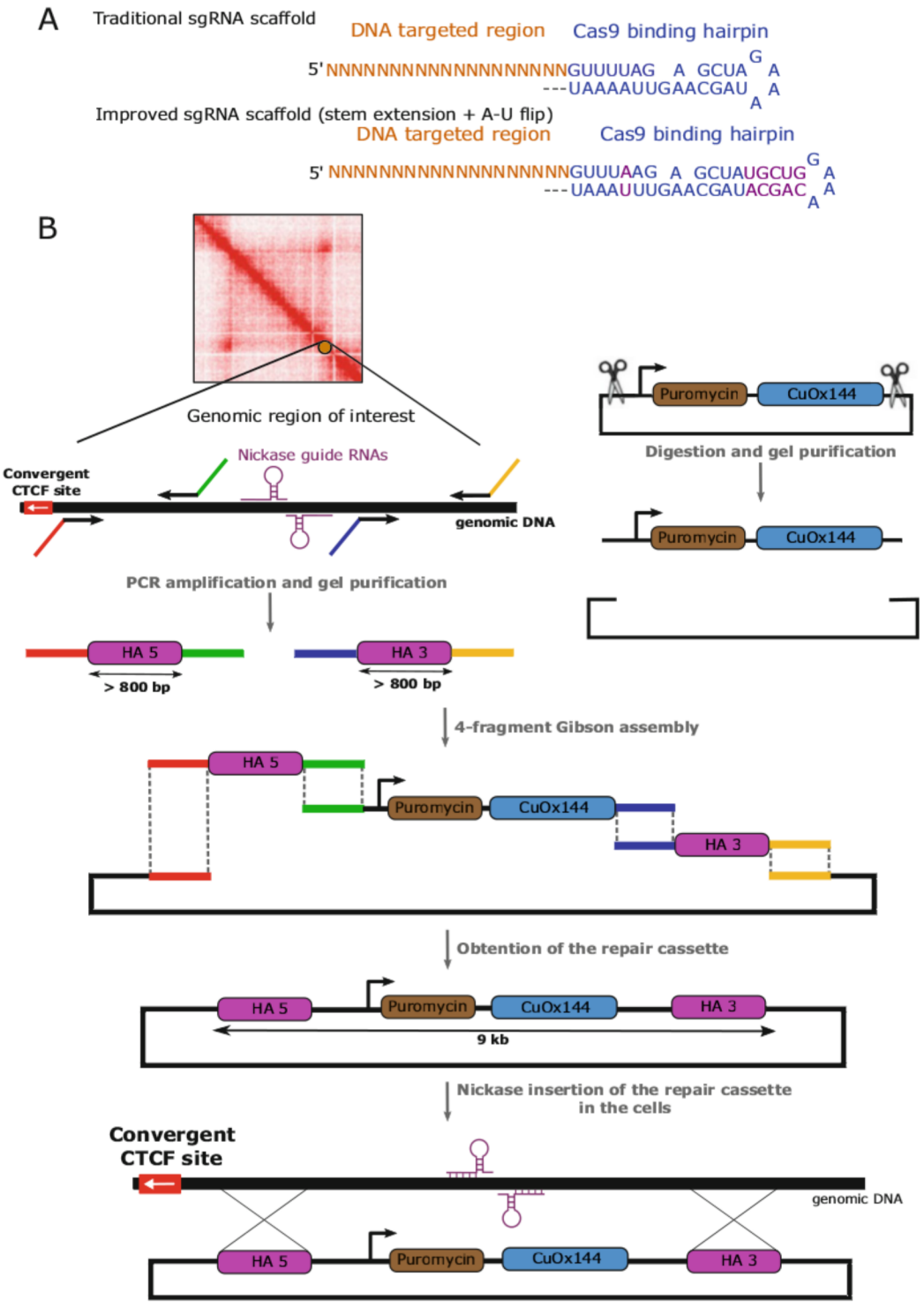


Fig. 1 Cloning method to generate repeat array repair cassettes at loop anchors. (**a**) Sequence of the traditional and optimized guide RNA scaffold designed in [12]. Modified nucleotides in the optimized scaffold are shown in purple. The dashes represent the *S. pyogenes* terminator. (**b**) Schematic representation of the cloning steps to generate the repeat array repair cassette. Top: Hi-C contact frequency map of a targeted

20. 10 μM forward and reverse primers to amplify homology arms (*see* Subheading 3.2 for design strategy).
21. HCT116 cells (American Type Culture Collection) expressing a fluorescently tagged CymR protein in the nucleus that will bind the CuO repeats (*see* **Note 1**).
22. Genomic DNA extraction kit (e.g., GenElute™ Mammalian Genomic DNA Miniprep Kit Protocol; Sigma-Aldrich).
23. 2 U/μL Phusion High-Fidelity DNA polymerase, supplied with 5X Phusion HF Buffer (ThermoFisher).
24. 10 mM dNTPs.
25. pDEST-CuOx144_Bxb1attB_loxP-PGKpuro-loxP containing the CuOx144 repeats and a puromycin-resistant gene, available on Addgene (#119903).
26. 20 U/μL NheI-HF (NEB).
27. 20 U/μL HindIII-HF (NEB).
28. Gibson assembly mix (NEBuilder HiFi DNA Assembly Master Mix; NEB).
29. Other restriction enzymes to verify Gibson assembly, will depend on the sequences of the cloned homology arms (*see* Subheading 3.3, **step 15**).
30. Cell culture medium: McCoy's medium supplemented with penicillin–streptomycin and 10% (v/v) fetal bovine serum (FBS).
31. pX335-U6-Chimeric_BB-CBh-hSpCas9-nickase_D10A which encodes the nickase D10A Cas9, available on Addgene (#42335).
32. JetPRIME buffer and reagent (Polyplus transfection).
33. 1× PBS (phosphate-buffered saline).
34. 0.5% trypsin–EDTA (diluted from 10× stock solution, Gibco).
35. Selection medium: Cell culture medium supplemented with 0.8 μg/mL puromycin.
36. Sterile cloning disks (diameter 3.2 mm) (Sigma-Aldrich).
37. 10 μM forward and reverse primers to screen array insertion into genome (*see* Subheading 3.5 for design strategy).
38. QuickExtract™ DNA Extraction Solution (Lucigen).

Fig. 1 (continued) chromatin loop. Long homology arms (>800 bp) are amplified by PCR on genomic DNA at the location of the loop anchor. The primers for homology arm amplification contain the overhangs (colored) for a Gibson assembly with the digested plasmid. Dotted gray lines show the overlap between the overhangs. The obtained repair cassette is then inserted in the cell genome by transfection of a nickase Cas9 together with two guide RNAs. Abbreviations: HA, homology arm

39. 5 U/μL GoTaq™ G2 Flexi DNA Polymerase (Promega), provided with 5× Green GoTaq Flexi Buffer and 25 mM $MgCl_2$ solution, or equivalent.
40. 4% (w/v) formaldehyde, made by diluting 16% formaldehyde (Thermofisher) in 1× PBS.
41. Vectashield antifade mounting medium with DAPI (LifeSpan Biosciences).
42. Widefield fluorescent microscope.
43. Glass-bottom petri dish suited for the chosen live-cell microscope (e.g., Petri fluorodish; Dutscher).
44. Live-imaging medium: antibleaching live cell visualization medium $DMEM^{gfp}$-2 (Evrogen), supplemented with 10% FBS, penicillin–streptomycin, and rutin (added fresh from 100× stock; Evrogen).
45. Fluorescent microscope setup for live imaging (see below).

3 Methods

3.1 sgRNA Design and Cloning

1. Use ChopChop (https://chopchop.cbu.uib.no/) to design the nickase sgRNAs targeting the region of interest (using the "CRISPR/Cas9 nickase" and "knock-in" options) and choose the highest ranked pair of guides which satisfies requirements (*see* **Note 2**).
2. Make the following modifications, if required, to the sgRNA oligonucleotide sequence design (*see* **Note 3** for an example).

   Do not include the NGG PAM sequence in the ordered oligonucleotides; On the side opposite to the removed NGG, add a G at the 5′ end if the oligonucleotide starts with a different nucleotide (and add a C at the 3′ end of the complementary oligo) for U6 transcription.

   Add cohesive ends identical to those of the digested vector containing the improved sgRNA scaffold: (a) before the 5′ end, on the side opposite to the NGG, add 5′ CACC on the oligonucleotide bearing the sgRNA sequence; (b) before the 5′ end, on the NGG side, add 5′ AAAC to the complementary oligonucleotide.
3. In a microtube, mix 5 μg improved sgRNA scaffold plasmid, 3 μL 10× BpiI Buffer G, 1 μL 10 U/μL BpiI and make the volume up to 30 μL with nuclease-free water. Incubate digestion reaction overnight at 37 °C.
4. Purify DNA with a PCR cleanup kit, following the manufacturer's instructions.

5. Measure DNA concentration with a NanoDrop spectrophotometer.
6. In a PCR tube, mix 1 μL 1 μg/μL forward sgRNA oligonucleotide, 1 μL 1 μg/μL reverse sgRNA oligonucleotide, 7 μL nuclease-free water, and 1 μL 10X BpiI Buffer G. Anneal the oligonucleotides with the following program in a thermal cycler.

   95 °C, 2 min;

   65 °C, 3 min;

   Cool to room temperature slowly (over 1 h).
7. In a microtube, mix 400 ng digested plasmid (from **step 4**), 2 μL 10 mM ATP, 2 μL 100 mM DTT, 2 μL 10× BpiI Buffer G, 1 μL 400 U/μL T4 DNA ligase, 1 μL 10 U/μL BpiI, and make the volume up to 20 μL with nuclease-free water. Add the annealed oligonucleotides (from **step 6**) to make a 30 μL ligation reaction volume and incubate overnight at 37 °C (*see* **Note 4**).
8. Add 0.5 μL 10 U/μL BpiI, 0.5 μL 10× BpiI Buffer G, and 4 μL nuclease-free water and incubate for 1 h at 37 °C.
9. Mix 2 μL of the ligation reaction with 50 μL of electrocompetent XL1-Blue bacteria and leave on ice for 10 min.
10. Electroporate the bacteria with the electroporator settings: 2.5 kV; 25 μF.
11. Add the electroporated bacteria to 700 μL SOB medium and incubate for 1 h at 37 °C.
12. Centrifuge for 3 min at 10,000 × *g*, room temperature, and discard the supernatant. Resuspend the bacteria in 100 μL LB and plate on ampicillin agar plate. Incubate overnight at 37 °C.
13. Pick four clones for each sgRNA with a sterile pipette tip. Make streaks of each clone on one ampicillin agar plate before putting the pipette tip in 5 mL of LB supplemented with 100 μg/mL ampicillin. Incubate the liquid culture overnight at 37 °C, 180 rpm, and the plate at 37 °C without shaking.
14. Extract plasmid DNA with a miniprep kit, following manufacturer's instructions.
15. In a microtube, mix 1 μg plasmid DNA, 3 μL 10× CutSmart buffer, 1 μL 20 U/μL BbsI-HF, and 1 μL 20 U/μL ScaI-HF, and make volume up to 30 μL with nuclease-free water. Incubate at 37 °C for 2 h.
16. Analyze digest on a 1% agarose gel. A unique 3196 bp linear band should be seen if the sgRNA sequence is incorporated; two bands are seen if the incorporation failed. Confirm sequence of digestion-verified clone by sequencing using the M13 forward primer.

17. Inoculate a colony of the verified clone (from the plate made at the same time as the miniprep culture in **step 13**) in 50 mL LB culture and incubate overnight at 37 °C, 180 rpm.
18. Purify plasmid DNA with a midiprep kit, following manufacturer's instructions.

3.2 Primer Design for Genomic PCR Amplification of the Homology Arms

1. Select a genomic region encompassing at least 800 bp starting just at the end of each of the two sgRNA sites (*see* **Note 5**). Note the homology arms' restriction maps for strategies to verify final Gibson assembly (*see* Subheading 3.3, **step 15**).
2. Design primers using the NCBI primer designing tool to specifically amplify the homology arms. Add the following 5′ overhangs for Gibson assembly (*see* **Note 6** and Fig. 1b).

   Forward primer for 5′ homology arm ("Red" in Fig. 1b): 5′-gagct cggtacgagtttacgtccagccaag.

   Reverse primer for 5′ homology arm ("Green"): 5′-ccggtag aatttcgacgacctgcagccaag.

   Forward primer for 3′ homology arm ("Blue"): 5′-aattcgagctcg aggtcgacggtatcgata.

   Reverse primer for 3′ homology arm ("Yellow"): 5′-ccctgccc gggctgcaggaattcgatatca.

3.3 Repair Cassette Cloning

1. Extract genomic DNA (gDNA) from 5×10^6 cells with the genomic DNA extraction kit, following the manufacturer's instructions. Use the exact same cell line that will undergo repeat array insertion.
2. In a PCR tube, mix 10 µL 5× Phusion HF buffer, 1 µL 10 mM dNTPs, 2.5 µL 10 µM forward primer, 2.5 µL 10 µM reverse primer, 60 ng genomic DNA, 0.5 µL 2 U/µL Phusion High Fidelity DNA polymerase, and make up the volume to 50 µL with nuclease-free water.
3. Run the following program on a thermal cycler (*see* **Note 7**).

   98 °C, 30 s;

   35×: [98 °C, 10 s; annealing temperature, 30 s; 72 °C, 30 s/kb product size];

   72 °C, 5 min;

   4 °C, hold.
4. Load the whole PCR reaction on a 0.8% agarose gel. Gel-purify the two homology arms using a gel purification kit, following the manufacturer's instructions (Fig. 1b).
5. In a microtube, mix 5 µg CuOx144 plasmid, 3 µL 10× CutSmart buffer, 0.5 µL 20 U/µL NheI-HF, and 0.5 µL HindIII-HF, and make up volume to 30 µL with nuclease-free water. Incubate overnight at 37 °C.

6. Load the whole digestion reaction on a 0.8% agarose gel. Gel-purify the two digestion fragments separately (6926 bp and 3199 bp) using a gel purification kit.
7. Assess DNA concentration of the PCR products and the digestion fragments with a NanoDrop spectrophotometer (*see* **Note 8**).
8. Perform the Gibson assembly with the two PCR amplified homology arms and the two digested plasmid fragments (4 fragments in total) (Fig. 1b). Use 0.2 pmol of total DNA fragments at a 1:1 vector–insert DNA molar ratio with 10 µL of Hifi Gibson assembly Master Mix in a 20 µL total reaction volume. Incubate the Gibson assembly reaction for 2 h at 50 °C in a thermocycler (*see* **Note 9**).
9. Mix 2 µL of the Gibson reaction with 50 µL of XL1-Blue bacteria and leave in ice for 10 min.
10. Electroporate the bacteria with the electroporator settings: 2.5 kV; 25 µF.
11. Transfer immediately the bacteria to 700 µL of SOB medium and allow for recovery 1 h at 30 °C, 180 rpm (*see* **Note 10**).
12. Centrifuge for 3 min at 10,000 × *g*, room temperature, discard supernatant, and resuspend bacteria in 100 µL LB.
13. Spread the bacteria on an ampicillin agar plate and incubate for 24 h at 30 °C (*see* **Note 10**).
14. Pick clones, grow them for 24 h in 5 mL LB medium at 30 °C, 180 rpm, and purify the plasmid with a miniprep kit, following the manufacturer's instructions (*see* **Note 10**).
15. Verify each plasmid by digesting 1 µg of DNA in a 30 µL reaction volume and analyze digested bands on a 1% agarose gel. Check for presence and orientation of homology arms and integrity of the full repeat array (*see* **Note 11**).
16. Make a 50 mL liquid LB culture of the verified clone, grow bacteria for 24 h at 30 °C, 180 rpm and perform the midiprep for high plasmid yield, following the manufacturer's instructions (*see* **Note 10**).
17. Verify the length of repeat array and correct insertion of the homology arms by restriction enzyme digestion (*see* **Note 11**).

3.4 Nickase-Mediated Knock-in

1. Seed 300,000 cells expressing a nuclear fluorescent CymR protein in a 6-well plate with 2 mL of complete McCoy's medium. Grow cells for 24 h in a humidified 37 °C incubator, 5% CO_2.
2. Mix 0.4 µg of nickase-expressing plasmid, 0.5 µg each of the two sgRNA plasmids and 0.6 µg of the repair cassette plasmid in 200 µL of jetPRIME buffer in a 1.5 mL tube and vortex.

3. Add 4 μL of jetPrime reagent, vortex and incubate for 10 min at room temperature.
4. Remove medium and add 2 mL of fresh medium to the cells. Add the transfection mix from **step 3** dropwise onto cells. Grow cells for 24 h.
5. Rinse cells with 2 mL of 1× PBS and add 300 μL 0.5% trypsin–EDTA to the cells. Incubate the plate for 3 min at 37 °C.
6. Prepare four different 10 cm plates with selection medium.
7. Dilute the detached cells in 10 mL of selection medium, and seed 200 μL, 500 μL, 1 mL, and the rest of the suspension into the different 10 cm plates (*see* **Note 12**).
8. Rinse cells with 5 mL of 1× PBS and add fresh selection medium twice a week until clones are clearly seen by eye (about 2 weeks).
9. Pick single colonies using cloning disks, put each of them in 24-well plates with selection medium and shake the plate vigorously to detach cells from the cloning disk (*see* **Note 13**).
10. Change the selection medium twice a week until cells reach 70% confluency.

3.5 Primer Design for Genomic DNA PCR Screening

1. For each side of the repair cassette (it is very important to check both sides to identify proper recombination events; *see* **Note 14**), select a genomic region outside of the homology arm (in the genome, as close as possible to the homology arm) and a region in the repair cassette that is close to the homology arm (Fig. 2a). The primers used inside the repair cassette we have used are.

   5′ side: 5′-ctaaagcgcatgctccagac

   3′ side: 5′-ttgtacaaagtggttgatgggg
2. Design PCR screening primers with the NCBI primer designing tool (*see* **Note 15**).

3.6 Clone Screening

1. Rinse cells with PBS and add 100 μL of 0.5% trypsin–EDTA. Incubate for 3 min at 37 °C.
2. Once the cells are detached, resuspend them in 600 μL of selection medium. Transfer 300 μL of the cells to a 1.5 mL tube and seed the other half in 1 mL of McCoy's medium in a 12-well plate for clone expansion.
3. Centrifuge the tubes for 5 min at 300 × *g*, room temperature. Discard supernatant and add 1 mL of PBS. Centrifuge for 5 min at 300 × *g*, room temperature and remove PBS.
4. Extract the DNA using the QuickExtract solution, following the manufacturer's instructions.

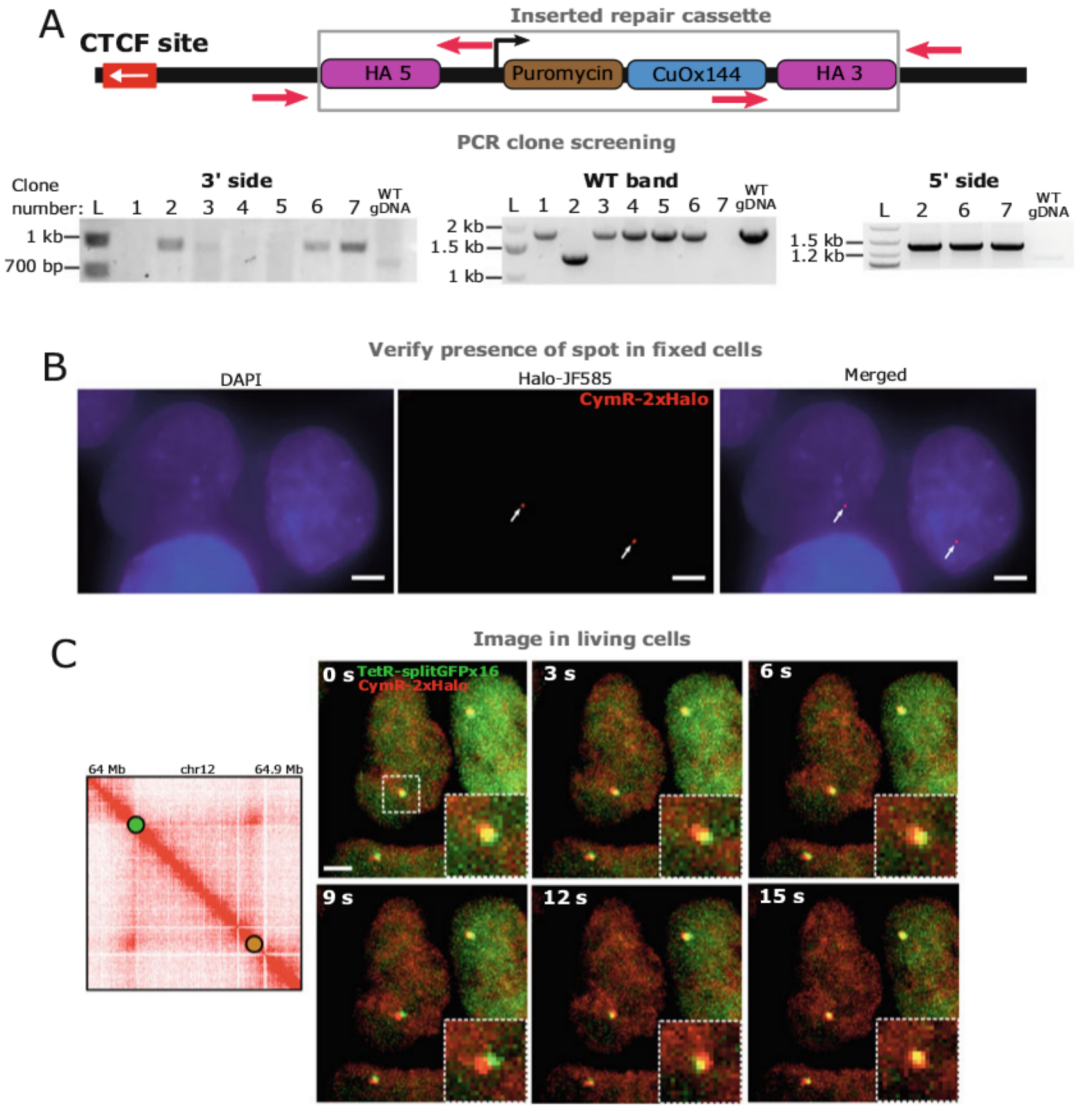


Fig. 2 Clone screening and imaging. (**a**) PCR screening to verify the insertion of the repair cassette. Top: Schematic of the inserted repair cassette in the genome. Pink arrows represent the location of screening primers at both sides of the repair cassette. Bottom: Agarose gel of 7 different clones after antibiotic selection. Left: Junction-PCR on the 3′ side (908 bp band is expected if the repair cassette is inserted). Middle: WT allele PCR (a band at 1.8 kb is expected if at least one allele is WT). Clone 7 has a homozygous insertion since no WT band is seen. Right: Junction-PCR on the 5′ side for clones that were positive on the 3′ side (a band at 1.4 kb is expected if the repair cassette is inserted). L: DNA ladder. (**b**) Fixed cell images of a PCR-screened positive clone. White arrows show the fluorescently tagged locus. Scale bar: 3 μm. (**c**) Live-imaging of loop anchors in HCT116 cells. Left: Hi-C contact frequency map of a 2 Mb region on chromosome 12 containing the two inserted operator arrays (circles) at the anchors of a 400 kb-long chromatin loop. Right: Spinning disk images of the loop anchors. Left anchor (green) was labelled by TetOx96 (inserted with the same cloning method described here for CuOx144) bound by TetR-splitGFPx16 and right anchor (red) was labeled by CuOx144 bound by a CymR-2xHalo (imaged with JF585). The inset shows a magnification of the region outlined by a white dotted line. The images are maximum intensity projections of 11 z-stacks each separated by 1 μm. One full z-stack was imaged every 3 s. Scale bar: 3 μm

5. In a PCR tube for each clonal gDNA preparation, mix 5 µL 5× Green GoTaq Flexi reaction buffer, 2.5 µL 25 mM $MgCl_2$, 0.5 µL 10 mM dNTPs, 0.5 µL 10 µM forward primer, 0.5 µL 10 µM reverse primer, 3 µL clonal gDNA, 12.88 µL nuclease-free water, 0.125 µL 5 U/µL GoTaq G2 Flexi DNA polymerase (*see* **Note 16**).
6. Run the following program on a thermocycler, with the annealing temperature determined using the Promega Tm calculator.

   94 °C, 2 min;

   35×: [94 °C, 30 s; annealing temperature, 30 s; 72 °C, 1 min/kb product size];

   72 °C, 10 min.
7. Analyze the PCR products on a 1% agarose gel (Fig. 2a).

3.7 Fixed-Cell Imaging for Verification of PCR-Positive Clones

1. From the 12-well plates containing PCR-positive clones, seed half of the cells in a 6-well plate for clone expansion and the other half on a glass coverslip inside a 6-well plate with 2 mL of selection medium.
2. 48 h after seeding the cells on the glass slide, rinse them once with 1× PBS and add 2 mL of 4% formaldehyde in 1× PBS and incubate for 15 min at room temperature.
3. Wash cells three times (1 min each) with 2 mL of 1× PBS at room temperature.
4. Put a drop of Vectashield + DAPI on a glass coverslip and transfer the slide with the cells onto the drop. Incubate for at least 2 h at 4 °C and protect from light.
5. Image the clones with a classic widefield microscope (acquire the whole cell height in z with each step separated by 0.3 µm) to verify the presence of one or two spots (depending on the zygosity of the insertion) in the nucleus (Fig. 2b).

3.8 Live-Cell Imaging of the Chromatin Loop Anchors

1. Seed 250,000 cells of a verified clone on a glass-bottom culture dish (suitable for the microscope to be used). Grow cells for 48 h.
2. Rinse the cells once with 1× PBS and change the medium to live-imaging medium.
3. Incubate for 2 h at 37 °C in the microscope chamber before imaging the living cells (*see* **Note 17**; Fig. 2c).

4 Notes

1. To increase the signal-to-noise ratio (SNR), and hence localization precision, of the visualized spot, it is preferable to sort the low fluorescence cells by FACS to limit background

intensity. We used an ePiggyBac CymR-NLS-2xHalo plasmid available on Addgene (#119907) to generate stable cell lines with low expression of the CymR protein for visualization of the integrated CuOx144 repeats. Diploid cell lines are preferred over polyploid ones.

2. Detailed and clear information on how to choose regions of interest are provided in [18]. Pick sgRNA sequences that are at least 50 bp away from repeated sequences so that the homology arm sequences closest to the DNA cut site are unique. The repeat array insertion sites should be located downstream of the last convergent CTCF site in order to not interfere with extruding cohesin but should be as close as possible to the loop anchor for precise quantification of the anchor-anchor distance. Insert the repeat array plasmid so that the antibiotic resistance transcription is directed outward of the TAD or chromatin loop to avoid interference with the loop extrusion process. Homology arms in euchromatin regions [17] and low percentage of repeated sequences will increase the insertion rates.
3. Sequence targeted by the sgRNA pair (in red the PAM sequence, in bold and underlined: sgRNA binding region).

AGTCCA**AGTTATGGAAGATCTCGATA**TCGAAGCCGTAGTCGTAGCAGAAGC

CGTAGCAGAAGCAACAGCA**GGAGTCGCAGTTACTCCCCA**AGGAG

sgRNA1: **TATCGAGATCTTCCATAACT**TGG

Oligonucleotides to order for sgRNA1 (in red: overhang to clone in U6 plasmid, in blue: on sgRNA side, in green: nucleotide added for U6 transcription).

5' CACCGTATCGAGATCTTCCATAACT

5' AAACAGTTATGGAAGATCTCGATAC

sgRNA2: **GGAGTCGCAGTTACTCCCCA**AGG

Oligonucleotides to order for sgRNA2 (in red: overhang to clone in U6 plasmid, in blue: on sgRNA side).

5' CACC**GGAGTCGCAGTTACTCCCCA**

5' AAACTGGGGAGTAACTGCGACTCC

4. Since sticky ends from BpiI sites are destroyed upon cloning the sgRNA sequence, only plasmids that incorporated the sgRNA sequence will not be digested by BpiI.

5. The beginning of the homology arm should be as close as possible to the NGG cut site (which should not be included in the homology arm) and no more than 100 bp away. The small sequence between the two sgRNAs will be deleted from the genome.
6. The overhangs should begin at the 3′ end of the cut NheI and HindIII sites on the CuOx144 plasmid. The enzymatic restriction site is destroyed after successful Gibson assembly since the full restriction site is not included in the overhangs. It is recommended to use 20 to 30 bp of Gibson overhang.
7. Use a high-fidelity DNA polymerase for the amplification of the homology arms. Use the ThermoFisher Tm Calculator to define the annealing temperature (using the "Phusion or Phire DNA polymerase" option). If unspecific bands are seen (higher annealing temperature is needed) or PCR amplification is not successful, run a gradient PCR with annealing temperatures separated by 3 °C each and choose the annealing temperature where the expected PCR product and no other bands are seen.
8. Ensure that plasmid and PCR fragment DNA concentrations are not lower than 25 ng/μL and that the A_{260}/A_{230} ratio is not lower than 1.5 for best Gibson assembly efficiency.
9. The Gibson assembly reaction should stay at 50 °C for at least 1 h. Increasing the incubation time leads to higher efficiency.
10. Repeat arrays can recombine in bacteria. It is therefore mandatory to grow them at 30 °C in recombinant-deficient bacteria to avoid losing a fraction of the repeats. Verify the length of the repeat array (5217 bp) at each step by restriction enzyme digestion (e.g., BamHI/ScaI digestion) to ensure full integrity of the size of the repeat array.
11. For restriction enzyme digestion verification of the plasmid, try to obtain one fragment containing the full repeat array to verify that no repeat was lost during bacteria growth. It is highly recommended to use restriction enzyme sites present inside the homology arms to verify their integration and orientation.
12. Using different dilutions of the cell solution ensures well-separated clonal colonies to pick after antibiotic selection.
13. The percentage of right insertions among resistant clones varies between 10–65% depending on the targeted locus. A few clones will not recover from the splitting. Picking 40 clones should ensure at least a few positive clones in low-efficiency conditions.
14. If using degenerated repeat cassettes, amplification of the whole array of repeats can be done to identify clones that did not lose some repeats during the recombination process. Optimization of the PCR conditions might be necessary.

15. Similar melting temperature and a GC clamp are preferable for the PCR screening primers.
16. Always add a PCR negative control with WT gDNA (no band is expected) to visualize potential unspecific bands. Once positive clones have been identified, perform a PCR to identify the WT allele and determine whether the insertion is heterozygous (presence of WT band) or homozygous (absence of WT band) before imaging the cells. Once a clone has been completely genotyped and verified by imaging, sequence the PCR products to verify correct recombination.
17. Detailed discussions of suitable live-cell imaging conditions can be found in [18, 19]. Briefly, organic dyes such as Janelia Fluor dyes (ligands of the Halo and Snap tags) [20] are preferable to fluorescent proteins due to their higher photostability, brightness, and versatility for multicolor imaging. Parameters to consider include the speed of acquisition or temporal resolution (especially for z-sectioning to avoid motion blur which decreases the localization accuracy), signal-to-noise ratio (SNR) and hence localization precision, total duration of imaging before photobleaching, and the ability of detecting two colors simultaneously if two separate structures are visualized.

References

1. Rao SSP, Huntley MH, Durand NC et al (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159:1665–1680
2. Cardozo Gizzi AM, Cattoni DI, Fiche J-B et al (2019) Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. Mol Cell 74(1):212–222.e5
3. Bintu B, Mateo LJ, Su J-H et al (2018) Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science 362:eaau1783
4. Boettiger A, Murphy S (2020) Advances in chromatin imaging at kilobase-scale resolution. Trends Genet 36(4):273–287
5. Rao SSP, Huang S-C, Glenn St Hilaire B et al (2017) Cohesin loss eliminates all loop domains. Cell 171:305–320.e24
6. Parmar JJ, Woringer M, Zimmer C (2019) How the genome folds: the biophysics of four-dimensional chromatin organization. Annu Rev Biophys 48:231–253
7. Fudenberg G, Abdennur N, Imakaev M et al (2017) Emerging evidence of chromosome folding by loop extrusion. Cold Spring Harb Symp Quant Biol 82:45–55
8. Davidson IF, Bauer B, Goetz D et al (2019) DNA loop extrusion by human cohesin. Science 366(6471):1338–1345
9. Sikorska N, Sexton T (2019) Defining functionally relevant spatial chromatin domains: it's a TAD complicated. J Mol Biol 432(3): 653–664
10. Hansen AS, Cattoglio C, Darzacq X et al (2018) Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus 9: 20–32
11. Beagan JA, Phillips-Cremins JE (2020) On the existence and functionality of topologically associating domains. Nat Genet 52:8–16
12. Chen B, Gilbert LA, Cimini BA et al (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155:1479–1491
13. Alexander JM, Guan J, Huang B et al (2018) Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. Elife 8:e41769
14. Mullick A, Xu Y, Warren R et al (2006) The cumate gene-switch: a system for regulated expression in mammalian cells. BMC Biotechnol 6:43
15. Germier T, Audibert S, Kocanova S et al (2018) Real-time imaging of specific genomic

loci in eukaryotic cells using the ANCHOR DNA labelling system. Methods 142:16–23

16. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823
17. Tasan I, Sustackova G, Zhang L et al (2018) CRISPR/Cas9-mediated knock-in of an optimized TetO repeat for live cell imaging of endogenous loci. Nucleic Acids Res 46:e100
18. Brandão HB, Gabriele M, Hansen AS (2021) Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. Curr Opin Cell Biol 70: 18–26
19. Schermelleh L, Ferrand A, Huser T et al (2019) Super-resolution microscopy demystified. Nat Cell Biol 21:72–84
20. Grimm JB, Muthusamy AK, Liang Y et al (2017) A general method to fine-tune fluorophores for live-cell and in vivo imaging. Nat Methods 14:987–994

2.4 Further genome editing improvements

We developed a molecular cloning strategy to efficiently generate CRISPR repair cassettes targeting numerous genomic loci and limiting as much as possible the tedious process of clone expansion and genotyping.

The final homology arms are obtained by a genomic PCR performed on the cell line used for CRISPR genome editing. This allowed us to capture potential Single Nucleotide Polymorphisms that are not represented in the averaged human genome.

After publishing this protocol[291], I further improved the design of the cassettes by (i) using a degenerated CuOx150 array and (ii) adding a Blasticidin-Thymydine Kinase (Bsd-TK) fusion. First, the degenerated CuOx150 array reduced the risk of losing repeats during the molecular cloning steps in bacteria or during DNA replication throughout cell divisions. Second, I was concerned that the strong transcription of the neomycin and blasticidin resistance genes used for clone selection would alter locus mobility or the loop extrusion process. Indeed, it has been proven that transcription confines chromatin motion[292] and that RNA Pol II are mobile barriers for cohesin[143]. Hence, transcription of these resistance genes could alter the motion of tracked loci and obscure the extrusion-dependent chromatin motion. To minimize the influence of the repeats on loop extrusion, I inserted them just outside of the CTCF anchors. Therefore, incoming cohesin molecules within the loop should not be affected by the repeats before reaching the CTCF anchors. To avoid the strong transcription emanating from the resistance genes, I therefore flanked the resistance cassettes with loxP and FRT sites to enable their removal using the Cre and flippase (Flp) recombinases, respectively. However, a negative selection marker was needed to ensure that growing clones have effectively performed the recombinations, rather than blindly genotyping all clones. TK gene expression selectively kills cells upon ganciclovir addition to the culture medium[293]. This suppressed the burden of expanding and genotyping clones that still expressed the Bsd antiobiotic resistance gene, which would not have been possible without the Bsd-TK fusion. Nonetheless, I could not design a working Neomycin-TK fusion (neomycin is used to select the CuOx150 array insertion) and had to rely on co-transfection events of Cre and Flp to select clones where both selection markers were removed.

Still based on the idea of reducing the burden of clone expansion and genotyping, I tested the use of the Diphteria Toxin (DT) gene. The expression of DT leads to cell death.

Positioning the DT gene outside of the repair cassette in the same plasmid enables to selectively kill the cells that integrated the whole plasmid or parts of the plasmid DNA that are not intended to be inserted. However, I did not detect a difference in correct insertion rate with or without the DT gene. Also, I was concerned that the DT gene expressed from the plasmid would tend to counter select transfected cells since plasmids can remain in the nucleus for several days. Therefore, I finally did not use the DT gene as a counter-selection gene. Clones with integrations of additional parts of the repair cassette plasmid were anyway discarded at the genotyping step.

This molecular cloning strategy allowed us to efficiently generate CRISPR repair cassettes targeting 14 different regions of the genome (7 pairs of chromatin loci). For reasons described below, only 5 of these cell lines could be obtained, due to either low CRISPR integration rate and/or excessively high fluorescent background in the parental cell line. A high variance of integration efficiency was observed between genomic loci: the percentage of genotyped clones with correct integrations ranged from 10 to 65% depending on the targeted locus. Additionally, the signal-to-noise ratio of the fluorescent spot was also variable as assessed by imaging, and called for further optimization.

2.5 Optimizing DNA visualization by maximizing signal-to-noise ratio and minimizing invasiveness

2.5.1 Tracking a spot within a fluorescent background

Our goal was to track in time a spot within a fluorescent background. The spot needs to display a higher fluorescent intensity than the background. This common challenge in imaging can be resolved by (i) increasing signal intensity and (ii) decreasing background intensity. Additionally, live-cell imaging requires to deal with photobleaching, the progressive loss of fluorescence intensity due to fluorophore excitation.

In our case, the signal intensity depended on binding of the fluorescently labelled TetR or CymR to TetO or CuO repeats, respectively. The background fluorescence arose from freely diffusing fluorescent TetR and CymR in the nucleus. I needed to amplify the signal of the fluorescent spot as compared to the background fluorescence. I detail below how I resolved the trade-off between maximizing signal-to-noise ratio and minimizing photobleaching.

2.5.2 Maximizing signal-to-noise ratio

Signal intensity directly depends on the number of repeats that are present in the array: the higher the number of repeats, the higher the spot intensity. I used 150 repetitions of the CuO array and 96 of the TetO array. I tried to limit the length of these arrays to increase CRISPR integration rate and minimize the potential artifactual effect of these arrays on native chromatin.

To limit background intensity, I FACS sorted the low fluorescently positive cells and filtered out the ones with a high fluorescence intensity after transposase-mediated integration of the freely diffusing TetR-GFP and CymR-2xHalo proteins in the genome.

The CuOx150 array was bound by a CymR-2xHalo fusion. Duplicating the Halo tag allowed us to double the fluorescent signal without increasing the number of repeats. The Halo-tag enables binding of organic fluorescent dyes, which are brighter and more photostable than fluorescent proteins. I used the recent Janelia Farm (JF) X646 dye, which is fluorogenic (does not fluoresce when not bound to its target), very bright and photostable[294].

In the green channel, background fluorescence intensity is higher than in far-red, even in absence of fluorophores, due to the presence of a higher level of cellular auto-fluorescence at this wavelength. To maximize and amplify the signal on the TetOx96 array, I used split GFP[295]. The 11 β-strands of the GFP protein constitute a β-barrel, which is split into two different parts. The 1-10 β-strands are expressed in the cell and freely diffuse without emitting fluorescence. The 11th β-strand is fused to the TetR protein and the whole GFP β-barrel can self-reconstitute (**Figure 16A-B**). To amplify the signal, GFP11 fragments were multimerized on the TetR fusion which harbored 16 GFP11 fragments so that each repeat should be bound by one TetR protein fused to 16 splitGFP (**Figure 16C**)[295]. In addition, some evidence suggested that the reconstitution of the GFP β-barrel is dynamic and can be reversed. The photobleached GFP1_10 fragments can thus be exchanged for intact fragments, hence reducing the photobleaching rate (**Figure 16D**)[295].

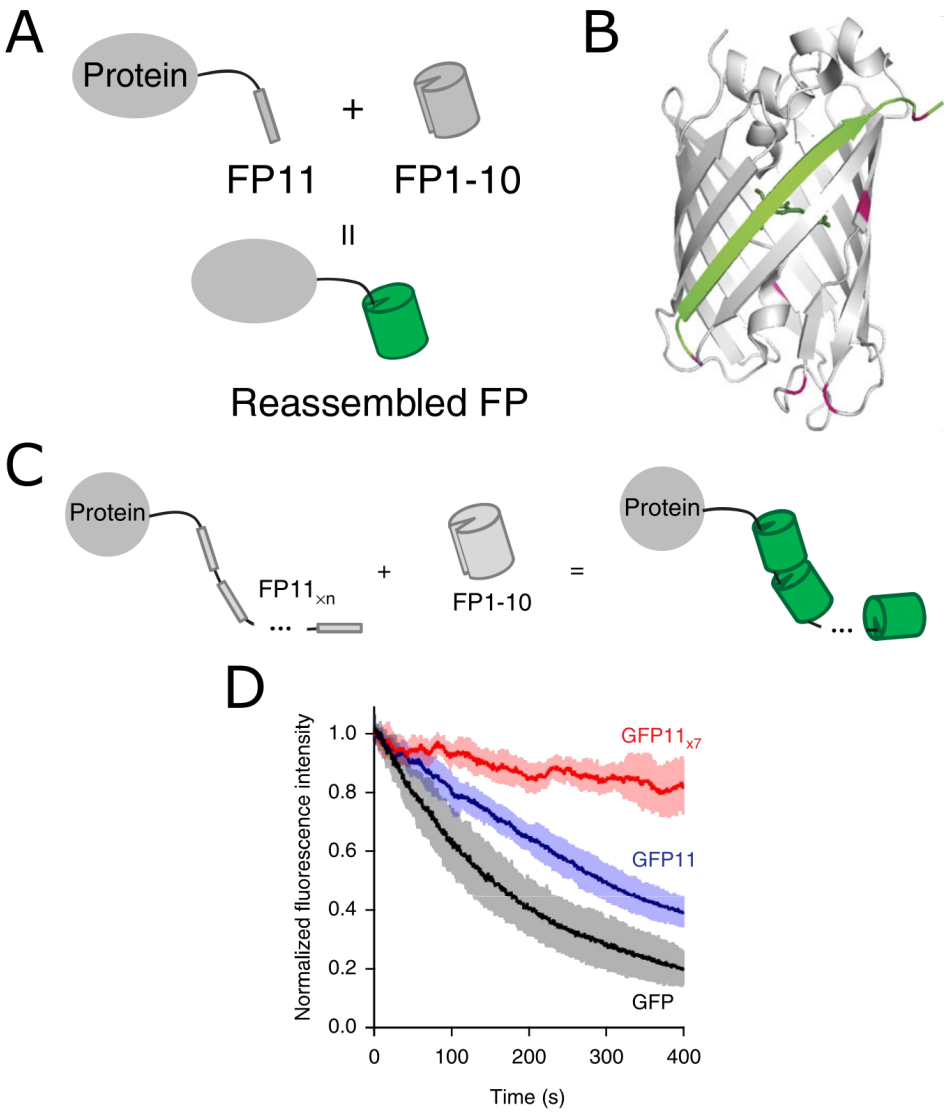


Figure 16: splitGFP allows signal amplification and long-term fluorescent imaging. A: A protein fused with the 11th β-strand of the fluorescent protein (FP) barrel can be fluorescently labelled by self-complementing with a freely diffusing FP1-10 containing 10 β-strands of the FP barrel. **B:** Schematic structure of the reconstituted GFP β-barrel. The green β-strand represents the 11th β-strand, which reconstitutes the whole FP. **C:** Signal amplification by multimerization. The protein of interest is labelled with several copies of the 11th β-strand. Hence, multiple reconstituted FPs label the protein of interest and increase its fluorescent signal as compared to the background. **D:** Fluorescent photobleaching time traces of GFP (black), splitGFP with a single GFP11 fragment (blue) and splitGFP with 7 repeats of the GFP11 fragment (red). Note that photobleaching rate is lower in the single GFP11 fragment (blue) as compared to the whole GFP (black). This suggests that replacement with unbleached fragments can occur at short timescales. Reproduced from [295,296].

We also optimized signal-to-noise ratio and reduced photobleaching using suitable optics. I used a spinning disk microscope allowing to optically section the sample, thus diminishing out-of-focus fluorescence and photobleaching. Moreover, I kept laser power and exposure time as low as possible to reduce photobleaching while keeping a sufficient signal intensity. The GFP and far-red dye could easily be separated using a dichroic mirror to avoid any crosstalk between the fluorescent channels. Thus, I optimized signal-to-noise ratio and reduced photobleaching using signal amplification techniques, photostable or exchangeable fluorophores, reducing background fluorescence intensity and choosing a suitable microscope (**Table 2**).

	Increase signal	Decrease background	Reduce photobleaching
CuOx150	- Bright organic dye - CymR-2xHalo fusion - High number of repeats	- FACS selection of cells with low background fluorescence - Fluorogenic dye	- Photostable organic dye
TetOx96	- split GFP multimerization - High number of repeats	- FACS selection of cells with low background fluorescence - GFP1_10 is dim when not bound to GFP11 fragment	- split GFP exchange
Optics	- Sensitive EMCCD cameras	- Optical sectioning - Low laser power - Low exposure time	- Optical sectioning - Low laser power - Low exposure time

Table 2 : Optimized parameters for maximizing signal-to-noise ratio and minimizing photobleaching during loop anchor tracking

2.5.3 Minimal number of TetO repeats enabling genomic locus visualization

We were concerned about the impact of the repeat length on chromatin biology. To minimize the invasiveness of our visualization method, I assessed the minimal number of repeats that allowed visualization of the repeat array.

Using the split GFP and the TetO array, I imaged TetO arrays of 24, 48 and 96 repeats. We found that 48 and 96 repeats allowed to detect the fluorescent spot reasonably well, while cells with 24 repeats displayed an extremely dim spot (**Figure 17**). Although minimizing

invasiveness would favor low number of repeats, I decided to keep the 96 repeats to obtain sufficient precision in spot localization and longer tracks.

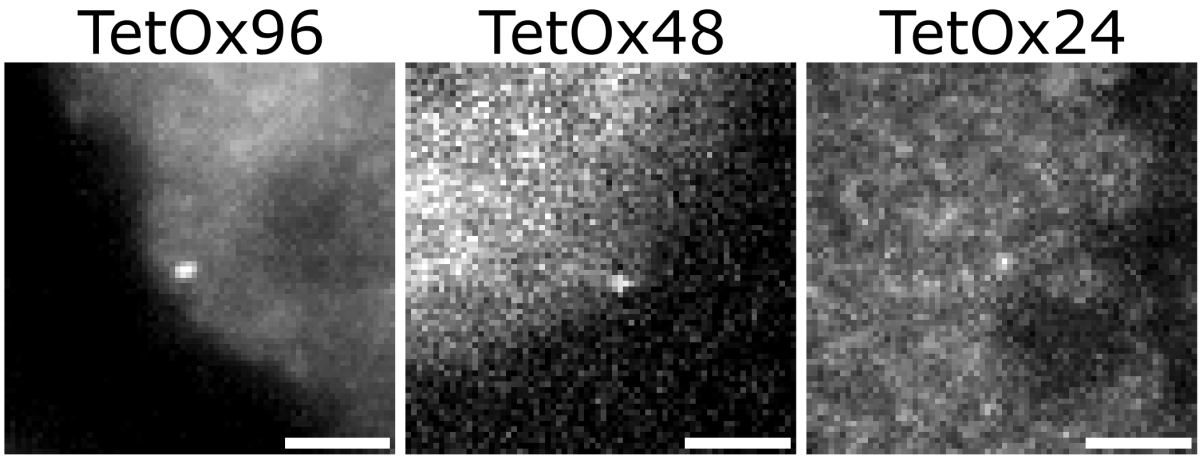


Figure 17: Images of spots with 96 (left), 48 (middle) or 24 (right) TetO repeats. TetOx48 and TetOx24 images were acquired using a Zeiss apotome to increase contrast. In these cells, the TetO repeats can barely be distinguished without the use of apotome. Scale bar: 2 µm.

2.5.4 Limitations and possible improvements

While the far-red channel (CuOx150 bound by CymR-2xHalo) had an extraordinary bright and photostable signal, we were limited in our acquisition by the GFP channel due to a relatively low signal. Theoretically, each repeat may be bound by 16 reconstituted GFP molecules, making the whole array bound by up to 1536 GFP molecules, which would result in an extremely bright spot. However, this was not observed, and several reasons can explain this limitation.

First, the TetO array contained only 96 repeats and increasing to 150 repeats would certainly improve the signal. Second, split GFPs are difficult to use because they require to obtain the correct ratio between the GFP1_10 and the TetR-GFP11x16 fragments to fully benefit from their potential and tuning this ratio is difficult. Theoretically, we would like to have 16 times more GFP1_10 than TetR-GFP11x16 molecules in the cell so that each TetR protein is fused to 16 fully reconstituted GFP molecules. By multiple lentiviral infections, I introduced the GFP1_10 fragments several times in the genome and noticed that this incomplete GFP already emitted low levels of fluorescence. Therefore, the background intensity might have been too high as compared to the signal intensity. Third, the GFP11

fragments were close to each other spatially, which may lead to steric hindrance (preventing all TetR-splitGFP molecules to be reconstituted at a given time) or fluorescent quenching between GFP. These issues might decrease the spot fluorescence intensity in the green channel.

Several exciting technical developments were published during my PhD, which could greatly enhance the imaging of the TetO array. New highly photostable and bright green fluorescent proteins, as compared to GFP, were identified in *Aequorea* species[297] and in *C. uchidae*[298]. These proteins, especially the StayGold[298], allow long-term tracking with bright signal and will certainly greatly help future live-cell imaging studies. Also, fluorogenic arrays of fluorescent proteins based on mutated proteins that reconstitute with low affinity were recently developed. Therefore, high exchange rate can occur during imaging, allowing photobleached molecules to be replaced by new fluorescent ones. They allowed virtually unlimited tracking of proteins and chromatin loci[299,300]. Signal amplification methods such as Suntag labeling[301] could replace our sub-optimal split GFP system. Another solution would have been to use the SNAP-tag (orthologous to the Halo-tag) to label the TetOx96 array and use bright organic dyes. However, I already labelled RAD21 with the SNAP-tag with the idea of tracking single cohesin molecules simultaneously to loop anchors. To reduce fluorescence background levels, I thought of using a Nuclear Localization Signal (NLS) fused to a Nuclear Export Signal (NES), which would have reduced TetR-GFP levels in the nucleus while still allowing some proteins to bind to TetO repeats in the nucleus[302].

Despite the frustration of using a sub-optimal split GFP system, I could nevertheless track loop anchors for a sufficiently long amount of time and with sufficient accuracy to extract valuable information.

In conclusion, I extensively tested and optimized the fluorescent visualization of chromatin loop anchors, as well as the CRISPR integration rate to enable the accurate quantification of chromatin loop dynamics on multiple endogenous loci.

3. Visualization and quantification of chromatin loop extrusion in living human cells

3.1 Cohesin-mediated loop extrusion creates dynamic loops in living human cells

I present below the current results obtained from the visualization and quantification of two endogenous chromatin loops in living human HCT116 cells. However, more data are currently being acquired, several analyses are still in development and finer analysis of the data will be performed in the future. Therefore, the results presented below should be considered as preliminary and are expected to be reinforced and expanded by new data and analyses.

Dynamics of cohesin-dependent loop extrusion in living human cells

Thomas Sabaté[1,2,3], Marie-Cécile Robert[2], Benoît Lelandais[1,4], Jean-Yves Tinevez[5], Edouard Bertrand[2,†], Christophe Zimmer[1,†]

1. Institut Pasteur, Université Paris Cité, CNRS UMR 3691, Imaging and Modeling Unit, F-75015 Paris, France.

2. IGH, University of Montpellier, CNRS UMR 9002, Montpellier, France.

3. Sorbonne Université, Collège doctoral, F-75005, Paris, France.

4. Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France.

5. Image Analysis Hub, Institut Pasteur, Université Paris Cité, Paris, France

† These authors contributed equally to the work.

Abstract

Animal genomes are folded into loops and Topologically Associating Domains by cohesin-mediated loop extrusion. This process has important consequences for gene regulation and other genomic functions, yet its temporal dynamics *in vivo* remain poorly known. Here, we dynamically tracked loop anchors to directly visualize and quantify cohesin-dependent loop extrusion at two genomic regions in living human cells. Using dedicated analyses methods, we found that anchors are brought in close proximity by extrusion on average 0.8-1.3 times per hour, and remain in closed loops for 2-16 min, corresponding to less than 27% of the time, on average. We found that both chromatin regions are rarely in an entirely open state as expected in absence of cohesin, but either closed or partially extruded. In addition, we estimated that cohesin molecule(s) effectively extrude DNA at a speed of 0.23-0.34 kb/s in living cells. Our results confirm that chromatin loops are formed by a dynamic process that generates short-lived rather than prolonged long-range chromatin interactions and provide novel quantitative estimates of loop lifetimes, frequency and extrusion speed in human cells. Our results have implications for a quantitative understanding of gene expression regulation or DNA repair mechanisms that exploit chromatin loop extrusion.

Introduction

Genome-wide chromosome conformation capture methods (such as Hi-C) revealed that mammalian genomes are folded into thousands or tens of thousands of 100-1500 kb large Topologically Associating Domains (TADs)[1–3]. TADs are defined by an enriched contact frequency within them, as compared to outside their boundaries, and hence appear as blocks on the diagonal of Hi-C contact matrices[1,2]. Their

boundaries are defined by convergently oriented CTCF sites, which act as barriers to extruding cohesin molecules[3–6]. TADs can display peaks of enhanced contact frequencies at their corners, which are interpreted as signatures of chromatin loops. TADs and loops critically depend on the cohesin complex, since degradation of its subunit RAD21 leads to disappearance of both from human cell lines[5,7]. Both TADs and loops, as defined from Hi-C data, are now understood as the result of cohesin-driven DNA loop extrusion, a process wherein the DNA is progressively pulled out of the cohesin complex, generating a loop of increasing size[5,8–11].

Cohesin-dependent loop extrusion is believed to play a role in key nuclear processes such as gene expression regulation, DNA repair and V(D)J recombination[12]. For instance, loop extrusion could favor long-range enhancer-promoter contacts within TADs and prevent ectopic gene expression across TADs[13–16]. Characterizing the dynamics of cohesin-mediated extrusion is therefore essential to better understand the mechanisms of these and other genomic functions.

Critical questions about loop extrusion include how frequently any given genomic region undergoes extrusion, how frequently loop anchors are brought into contact, how long anchors remain in contact, and how rapidly DNA is extruded by the cohesin complex. Answers to these questions remain partial[5,10,11,17,18].

Experimentally, TADs and loops are usually characterized using contact maps that represent statistical averages over millions of cells in bulk Hi-C experiments. However, single-cell Hi-C[19], multiplexed DNA FISH studies[20–22], single-particle tracking[23] and polymer simulations[9] revealed that chromatin structure can vary from cell to cell within an isogenic population. Furthermore, Hi-C and DNA FISH are intrinsically limited to fixed cells, and therefore ill-suited to characterizing the

dynamics of specific chromatin regions. Thus, a full understanding of loop extrusion requires analyzing this process in single and living cells.

Experiments that recapitulate loop extrusion *in vitro* have shed light on some dynamic features of loop extrusion, *e.g.* by measuring the cohesin-mediated extrusion rate of 0.5-1 kb/s[10,11]. However, it remains uncertain whether extrusion occurs at a similar, higher or lower rate in chromatin within living cells, where many molecular factors may potentially slow down or on the contrary accelerate this process.

Two recent studies used live cell imaging to analyze chromatin loop dynamics at a single genomic locus each in mouse embryonic stem cells (mESCs). These studies confirmed that TADs are highly dynamic structures, and estimated that loop anchors are in close contact 3-6%[17] or 20-30%[18] of the time for median durations of 5-30 min[17,18]. One study showed that the TAD was found to be 92% of the time in a partially extruded state[17], consistent with the estimation that several cohesin molecules actively extruded TADs at the same time[17,18]. However, these results were limited to a single locus each in mESCs, whose chromatin is less condensed and features many differences with chromatin in differentiated cells[24–28]. In addition, these studies did not provide measures of loop extrusion speed. Thus, a full dynamic characterization of loop extrusion in single human cells remains lacking.

Here, we used live-cell fluorescence microscopy together with quantitative analysis methods validated on polymer simulations to visualize and quantify loop extrusion by labelling several pairs of endogenous loop anchors in living human HCT116 cells. We found that the looped state was occasional (<26%), transient (2-16 min) and infrequent (it occurred once per hour). Moreover, loops were found to be submitted

to almost constant extrusion and were rarely found in their purely relaxed state. Finally, we estimated that loops are effectively extruded at a rate of 0.23-0.34 kb/s *in vivo*.

Results

Visualizing chromatin loop anchor dynamics

To visualize and quantify cohesin-mediated chromatin looping in living cells, we labelled endogenous loop anchors by CRISPR-mediated insertion of TetOx96 and CuOx150 arrays, bound by TetR-splitGFPx16 and CymR-2xHalo (imaged with the bright and photostable JFX646 dye[29]) proteins, respectively (**Fig. 1A**). Since transcription is known to constrain chromatin motion and interfere with loop extrusion[30], we removed the antibiotic resistance genes used to select array integration from the genome by Cre and Flippase-mediated recombinations (**Fig. 1A**).

Chromatin loops or TADs exhibit different genomic sizes and are characterized by different contact patterns in Hi-C maps[31,32]. To account for this variability, we selected two pairs of anchor loci, Lv1 and Lv4, separated by genomic distances of 350 kb and 570 kb, respectively, and both corresponding to a TAD with corner peak. We also selected one pair of anchors, Tv1, separated by 925 kb and corresponding to a TAD without corner peak (**Fig. 1C**). In order to minimize the complicating effects of transcription on chromatin motion and long-range contacts[30,33,34], the domains were selected not to contain genes at their anchors and to have no or low gene expression within the domain. All three domains exhibited SMC1, RAD21 and CTCF ChIP-Seq peaks at both anchors, as well at least one pair of convergent and bound

CTCF sites. In addition to these three genomic loci, we generated two control cell lines. In the first control cell line, the genomic distance between labeled loci was 581 kb, similar to Lv4, but one of the two labeled loci was located outside of a TAD and not at a CTCF border, therefore extrusion is not expected to induce prolonged contacts between the two loci. Thus, this cell line effectively served as a negative control of CTCF-dependent loops (**Fig. 1C**). In the second control cell line, we inserted the two fluorescent reporters close to each other, at a mid-array distance of 6 kb, to serve as a proxy for loops in the closed state, *i.e.* when the two loop anchors are in close proximity. This cell line thus acted as a positive control for closed states (**Fig. 1C**).

We used spinning disk confocal microscopy to image the cells in 3D every 30 s during 2 hours preceded or not by a 2-hour auxin treatment (**Fig. 1B**, **Fig. S1**). Then, we detected and tracked fluorescent spots, and computed their relative 3D distance as function of time. We obtained 154-606 distance *vs* time tracks per cell line and experimental condition (**Fig. 1D-F**, **Table S1**). In order to avoid measuring sister-chromatid interactions, we removed replicated spots from the analysis and therefore only analyzed cells in G1 or early S-phase (**Fig. S2**).

For each cell line, excepting the closed control, we either left cells untreated or treated them for 2 hours with auxin, leading to depletion of RAD21, a condition in which no loops are expected[5] (**Fig. S3**). We used the AtAFB2 auxin-dependent degron, which yielded an efficient depletion (91% depletion after 1 hour of auxin treatment, **Fig. S3B-C**) with minimal basal degradation (>76% of endogenous RAD21 level in untreated cells, **Fig. S3D,F**)[35]. Because the presence of cohesin-mediated loop extrusion is expected to reduce anchor-anchor distance, auxin treatment is expected to increase this distance, on average. This is indeed the case

for all three domains Lv1, Lv4 and Tv1 (**Fig. S4A**). Likewise, we observed an increase in chromatin motion upon RAD21 depletion, in agreement with predictions from polymer simulations and prior observations[17,18] (**Fig. S5**).

Thus, we could visualize the dynamics of endogenous chromatin loci in living human cells in presence or absence of loop extrusion and accurately compute the 3D distance between loop anchors as function of time (**Table S1**).

The closed state is infrequent and transient

At any given time, a pair of anchors is in only one of three possible states: (i) the open state, where the chromatin region between anchors is free from cohesin, (ii) the extruding state, where one or more cohesin molecules actively extrude one or several loops and (iii) the closed state, where the chromatin between anchors is fully extruded and the anchors are maintained in close proximity by cohesin (**Fig. 1B**). We first aimed to quantify the fraction of time, frequency and lifetime of closed states.

The Lv1 and No loop cell lines have not been analyzed yet, therefore, we present and discuss results from the Lv4 and Tv1 cell lines. As already observed in previous studies[17,18], our control for the closed state had an unexpectedly large distribution of distances. Simulated polymers in the looped state with localization errors higher than the ones measured on this cell line exhibited lower 3D distances than the experimental measured distances, suggesting highly decondensed DNA in a fraction of the cells (**Fig. S4B**). This closed control cell line could not be fully genotyped and might exhibit unexpected genomic rearrangements. A different cell clone will be imaged and analyzed to check if this behavior is reproducible and better understand the unexpected behavior of this control.

Below, we discuss results obtained on Lv4 and Tv1 cell lines. We first segmented tracks into intervals of closed states (without differentiating the open and extruding states), using a method previously validated on polymer simulations[36]. This method involves two thresholds: a spatial threshold defined using closed control distances and a temporal threshold, defined on the RAD21-depleted cells. We thus identified temporally sustained low distance intervals, which are absent upon cohesin depletion, as a proxy for closed state intervals (**Fig. 2A, Fig. S6**). Specifically, we set the thresholds to obtain a closed state fraction of 5% in RAD21-depleted cells (proxy for false positive detection rate) (**Fig. S7D-F**), and used them to analyze tracks from untreated cells. This led to a mean closed state fraction of 26 ± 2% for Lv4 and 18 ± 2% for Tv1 (**Fig. 2B**).

In addition to these fractions, our segmentation of tracks allowed us to estimate the frequency of loop closure, *i.e.* the number of transitions from extruding to closed states per unit time. We found that closed states occurred at a rate of, on average, 1.32 ± 0.17 and 0.78 ± 0.14 times per hour, for Lv4 and Tv1 respectively (**Fig. 2C**), implying that during a typical G1 phase of 10 hours, the closed state was formed between 8 and 13 times. By exponential fitting of the distribution of closed state durations, we estimated that closed state median lifetimes were 9.3 ± 0.69 min for Lv4 and 11.6 ± 1.1 min for Tv1 (**Fig. 2D**).

Interestingly, we found that fraction and frequency of closed states were higher in the smaller and stronger loop Lv4 than in the larger and weakier TAD Tv1 (**Fig. 2B-D**). Hi-C patterns and more specifically the appearance of a ‘dot’ at TAD corners might be explained by a higher probability of closed state. Assuming constant processivity and extrusion speed, the shorter loop Lv4 should close more frequently than the longer loop Tv1 and have higher fractions and lifetimes of closed states. Our results

are consistent with this prediction (**Fig. 2B-C**). However, assuming homogeneous loading and unloading rate of cohesin to and from chromatin throughout the genome, larger TADs should be bound by a higher number of cohesin molecules. It is therefore not *a priori* evident whether larger loops should exhibit lower or higher fractions of closed states than smaller loops. Using modeling and the different targeted loops, we plan to further analyze the interplay of cohesin processivity and cohesin density and its role on loop dynamics.

Upon auxin treatment, as expected, the fraction, frequency and lifetime of closed states always exhibited the lowest values (**Fig. 2B-D**), while the highest values were found in the closed state control (**Fig. S7A-C**). However, it was not purely identified as the closed state due to its unexpected large distribution of distances.

Using different pairs of spatial and temporal thresholds, while still assuming 5% of false positive rate, we found closed state fractions of 21-27% and median lifetimes of 3-13 min for Lv4, 11-20% closed state and median lifetimes of 3-12 min for Tv1 (**Fig. S7D-F**). These results mildly changed when using a false positive rate of 1%. Under these more conservative conditions, we found closed state fractions of 7-10% for Lv4 and 8-12% for Tv1 (**Fig. S7D**) and median lifetimes of 2-16 min for Lv4 and 3-12 min for Tv1 (**Fig. S7E**). Although median lifetimes slightly depended on the chosen temporal threshold (**Fig. S7F**), their values were consistently within the 2-16 min range.

Thus, we found that the closed states occur occasionally and transiently in human HCT116 cells and that loop anchors contact each other about 8-13 times during a 10-hour G1 phase, for a total summed duration of about 2 hours.

Chromatin loops are extruded constantly

After quantifying the closed states, we aimed to quantitatively characterize the open and extruding states.

First, we estimated the average fraction of the TAD length that was unextruded (*i.e.* the portion of the TAD left without cohesin-dependent extrusion) by comparing the mean squared distances in untreated *vs* auxin-treated cells: $\frac{R^2}{R_0^2}$, where R_0^2, is the mean squared distance in auxin-treated cells and R^2 is the mean squared distance in untreated cells. Using this method, we observed that 24% and 36% of the Lv4 loop and Tv1 TAD, respectively, were left unextruded on average. Thus, by subtraction, we found that extrusion and CTCF sites together reduced the physical distance between the two fluorescent reporters by 76% and 64% for Lv4 and Tv1, respectively (**Fig. 2E**). Since a major part of the TAD is extruded on average, this suggested that TADs were generally under active extrusion, rather than free from cohesin molecules.

We next used a method previously validated on simulations to quantitatively estimate the fractions of each loop state (closed, extruding and open) in each cell line from the total distribution of anchor-anchor distances. This method requires knowing the anchor-anchor distance distribution in the closed and open states. For the former distribution, we used the distances in untreated cells during time intervals segmented as closed states, whereas the latter distribution was determined from all distances measured in RAD21-depleted cells). From these two distributions, we can model the extruding state and retrieve the fractions of each loop state[36].

Unsurprisingly, the closed control exhibited 100% of closed states in both cell lines. In the auxin-treated cells, we estimated 2% and 9% of median closed states for Lv4

and Tv1, respectively, which is close to the chosen threshold for closed state segmentation that yields 5% false positive in the dynamic analysis (see above). More significantly, we estimated open state fractions of 87% for Lv4 and 71% for Tv1 (**Fig. 2F**), consistent with the disappearance of most chromatin loops expected from cohesin depletion. In the untreated cells, by contrast, we estimated 24% of closed states for Lv4, similarly to the temporal analysis (**Fig. 2B**) and 40% for Tv1 (**Fig. 2F**), more than double the fraction obtained from the temporal analysis (**Fig. 2B**). We further estimated extruding state fractions of 82% and 54% for Lv4 and Tv1, respectively. Strikingly, the open state was completely absent (0%) in both cell lines (**Fig. 2F**).

Thus, the extruding state is the most frequently observed state in both cell lines and our data indicate that loops are never found in a completely relaxed (open) state. Therefore, our results suggest that TADs are constantly undergoing active cohesin-dependent DNA loop extrusion.

Cohesin molecules effectively extrude loops at 0.3 kb/s in living cells

Next, we aimed at determining the speed of loop extrusion *in vivo*, which has not hitherto been directly measured in living cells. This task is challenging, because the progressive decrease in anchor-anchor distance expected from extrusion can be obscured by stochastic fluctuations due to random chromatin motion compounded by random localization errors[36]. To reduce these stochastic fluctuations, we reasoned that timepoints immediately preceding the closed loop states should all be undergoing active extrusion and therefore averaged the distance time series from a

large number of cells, after alignment on the starting time of the closed states[36] (**Fig. 3A-B**). We filtered out closed states preceded by another closed state to ensure that timepoints before the closed state were in the extruding state. We then computed a linear fit to the ensemble mean squared anchor-anchor distance over time intervals before closed states, allowing us to estimate the effective speed at which loops are extruded *in vivo*[36].

By definition, anchor-anchor distances preceding closed states are larger than during closed states, potentially leading to a biased estimation of extrusion speed. To verify that our method does not lead to erroneous estimates in absence of extrusion, we randomly shuffled all time points within single time series, thereby destroying any signature of processive dynamics and providing a negative control for extrusion speed estimation. Segmentation of closed states, alignment and fitting of these randomized time series yielded loop extrusion rates of 0.03 ± 0.05 kb/s and -0.03 ± 0.06 kb/s for Lv4 and Tv1, respectively, consistent with an absence of extrusion (**Fig. 3B-C**). Therefore, both our closed state segmentation and extrusion speed fitting methods did not bias the extrusion speed estimation in time series where extrusion is not expected to be detected.

We next proceeded to estimate extrusion speed in untreated cells. From n=427 and n=606 single time series for Lv4 and Tv1, respectively, we averaged ~338 and ~319 closed states for Lv4 and Tv1, respectively. By linearly fitting the ensemble mean squared anchor-anchor distance, we estimated effective loop extrusion speeds of 0.23 ± 0.07 kb/s for Lv4 and 0.34 ± 0.11 kb/s for Tv1 (**Fig. 3C**). The two targeted regions exhibited statistically significant extrusion speeds, which may indicate that cohesin molecules do not effectively extrude at similar speeds on different locations

of the genome, as previously observed[5]. These estimates constitute the first direct estimate of effective extrusion speed in living cells.

Discussion

By tracking chromatin loop anchors in living human cells, we found that contacts between loop anchors are observed 7% to 27% of the time, occur on average 0.8 to 1.3 times per hour and last for 2-16 min. This means that anchor-anchor contacts are short compared to the typical 10-hour G1 phase and occur on average 8-13 times during a single G1 phase. In addition, we found that both examined genomic loci (of 570 and 925 kb) were almost constantly submitted to loop extrusion and were never found in a purely open state 0% in both cell lines. Moreover, we estimated that DNA is extruded effectively at rates of ~0.3 kb/s in living cells, and that this rate may vary between different regions of the genome.

Our estimates of the duration of closed loop states are in relatively good agreement with two previous studies in mESCs, despite the different mammalian models and different analysis methods. Both studies found loops to be dynamic and short-lived, with median lifetimes of 3-30 min[17,18] similarly to our 2-16 min estimate in human cells. The fraction of time spent in the closed state was more variable between studies and ranged from 3-6% for an endogenous 505 kb loop[17] to 27% for a strong synthetic 150 kb loop[18], compared to 7-27% and 8-18% in our endogenous 570 and 925 kb loops in human cells. A high number of confounding factors (human cancer vs mESCs cells, heterogenous cohesin dynamics across the genome, genomic distance and loop insulation) may explain the remaining differences and more work

is needed to determine the biological *vs* technical causes of differences between these results.

Our finding that 570 kb and 925 kb sized chromatin loops are most often found in a partially extruded state and rarely in an open state, similarly to a previous study[17], argues in favor of a model where loops emerge from a collection of growing loops[37]. Therefore, it is likely that several cohesin molecules are simultaneously bound within the region of a single loop.

Our study provides the first direct estimate of DNA loop extrusion speed *in vivo*. Because several cellular factors could potentially accelerate or decelerate extrusion rates in cells as compared to *in vitro*, our estimated speed of 0.23-0.34 kb/s is surprisingly close to estimates of 0.5-1 kb/s from *in vitro* experiments with purified proteins[10,11]. We note that while *in vitro* studies measured the speed of single cohesin molecules, we measured an effective loop extrusion speed *in vivo, i.e.* the speed at which loops are extruded, independently of the number of cohesin complexes extruding the loop. This estimate potentially reflects the combined effect of occasional stalling of single molecules and possibly of the action of multiple cohesin complexes (monomeric or dimeric) extruding simultaneously. Because of stalling, single cohesin complexes may potentially reach higher peak speeds, but because of the possibility of multiple cohesin complexes, the average speed of a single complex may be lower than our effective extrusion speed estimate. Our estimated effective speed of 0.3 kb/s is also remarkably consistent with the speed of about 0.38 kb/s inferred indirectly from a dynamic Hi-C study at distinct time points[5]. Thus, as compared to other nuclear motors (*e.g.* RNA Polymerase II transcribing at about 0.02 kb/s[38], 0.025 kb/s for the RSC translocase[39]), cohesin is a fast motor

extruding DNA at high rates. This elevated speed likely contributes to the transiency and highly dynamic nature of cohesin-induced long-range interactions.

We found that loop extrusion speed was statistically different between the two considered regions. While it is yet not clear whether this difference emerges from a technical or biological difference, it may indicate that cohesin can effectively extrude at various rates in different regions of the genome. Indeed, after auxin-dependent and auxin washout to allow for a resynthesis of cohesin, it was shown that the reappearance of loops was not synchronized throughout the genome[5]. For instance, loops in the B compartment reappeared at a slower rate than in the A compartment[5]. Although this might not come from a difference in the speed of single cohesin molecules, this might emerge from a different binding rate of cohesin or CTCF between these two regions[40].

We note that several unknowns impose limitations to our analysis. First, the high effective rate (~0.3 kb/s) at which chromatin loops are extruded could prevent extruded chromatin to reach equilibrium. If extrusion is not at equilibrium, out of equilibrium polymer models might be needed to better describe the loop extrusion process. Second, our analysis ignored cohesin stalling, which was observed *in vitro* when cohesin encounters roadblocks[41]. Since the TADs chosen here also contained bound CTCF sites, cohesin may be susceptible to stall within the loop. A clear understanding of how CTCF binding site affinity and number of CTCF sites affect cohesin pausing could help to provide better estimates. Third, as previously assessed in polymer simulations, the finite localization precision of fluorescent spots affects closed state segmentation and distance measurements, which may lead to an underestimation of loop extrusion speed[36].

Taken together our results describe the highly dynamic nature of cohesin-induced interactions in human genomes. They support a model where cohesin complexes almost constantly extrude loops at high rates, but yield transient rather than prolonged contacts between loop anchors. This model has implications on our quantitative understanding of how frequent but short-lived contacts produced by extrusion between enhancers and promoters may help regulate gene activation. The generation of short-lived long-range interactions by loop extrusion is likely to be exploited by other biological processes that involve direct contacts between DNA sequences, such as DNA repair, which may also tune and alter loop extrusion dynamics[17,42].

Acknowledgements

We acknowledge the MRI imaging facility, part of the national infrastructure France-BioImaging supported by the French Nation Research Agency (ANR-10-INBS-04, Investments for the future). T.S. was supported by a Contrat Doctoral Spécifique aux Normaliens and Fondation ARC pour la recherche sur le cancer. We also acknowledge Investissement d'Avenir grant ANR-16-CONV-0005 for funding computing resources used in this work.

Author Contributions

T.S. designed the project, performed the experiments and the analysis and wrote the manuscript. M.C.R. participated in cell line construction. B.L. developed the image pre-processing pipeline and the extrusion speed analysis. J.Y.T. contributed to image analysis. C.Z. and E.B. supervised the project and wrote the manuscript.

Competing interests

All authors declare no competing interests.

Materials and methods

Cell line culture, generation and treatment conditions

Cell culture

HCT116 cells were cultured in McCoy's medium supplemented with GlutaMAX (Thermo Fisher Scientific 36600021), 10% Fetal Bovine Serum (FBS) and Penicillin-Streptomycin (50 U/mL and 50 µg/mL respectively). Cells were grown at 37°C in a humidified incubator with 5% CO_2. Cells were tested monthly for the presence of mycoplasma.

Cell line generation and genome-editing

Genome-editing was performed using a nickase Cas9. We co-transfected a repair plasmid, a nickase Cas9 expressing plasmid and a pair of single guide RNAs (sgRNAs) (total of 2 µg of DNA) using Prime Jet according to manufacturer's protocol. The pair of sgRNAs were designed using ChopChop[43].

300,000 cells were seeded in a 6-well plate and transfected 24 hours later. Less than 24 h after transfection, cells were detached from the well and split into four different 10 cm plates for selection. Each 10 cm plate contained a different dilution of the initial 6-well plate (from 1/40 to 4/5). Cells were kept under selection until unique colonies were seen (about 2 weeks). Single colonies were then picked using cloning disks and put into 24-well plates. Once sufficiently grown, each clone was split in half: One half was used for clone expansion and the other half was seeded on a glass slide for image-based screening. Three days after seeding, Halo tag was

labeled, if needed, with JFX646 by incubating the cells for 15 min at 37°C. Cells were then fixed with 4% formaldehyde (28908) in PBS for 20 min and slides were mounted in Vectashield antifade medium with DAPI (Vector H-1200-10). Clones were imaged with a widefield microscope (Zeiss Axioimager, 63X Plan Apochromat NA=1.4 objective with a LED Xcite 120LED as illumination source and an ORCA-Flash4 LT Hamamatsu camera with 2048x2048 pixels and a pixel size of 6.5 µm). Clones that displayed one or two fluorescent spots per nucleus were further split in a 6-well plate and genomic DNA (gDNA) was extracted (Lucigen). PCR genotyping was made by amplifying 5' and 3' junctions. The unmodified wild-type band was amplified to assess the zygosity of the insertion. High-quality gDNA of clones verified by images and PCR was purified (Promega A1120) and each PCR fragment was sent for sequencing. Finally, we checked the TetOx96 array integrity by PCR amplifying and sequencing the whole array with the 3' junction primer.

We first homozygously inserted the RAD21-mAID-SNAP-IRES-Hygromycin fusion, using 100 µg/mL hygromycin for selection. Next, we inserted the AtAFB2-IRES-Puromycin fusion at the AAVS1 locus by selecting cells with 1 µg/mL puromycin (Invivogen ant-pr-1) to enable auxin-mediated RAD21 degradation. We chose the AtAFB2 degron protein instead of the more common OsTIR1 because AtAFB2 was reported to minimize basal degradation of the degron-tagged protein[35].

We then expressed the fluorescent reporters needed to visualize the repeat arrays. Using piggybac transposase, we inserted CymR-2xHalo and TetR-GFP11x16 in the cells. We infected the cells with lentiviruses containing the monomeric GFP1-10 fragment to reconstitute the split GFP. Cells were sorted three times once a week to keep only low expressing levels of the reporter proteins. Then, we re-infected the cells with the GFP1_10 lentiviruses to increase the GFP1-10 over TetR-GFP11x16

ratio and optimize the signal from the multimerized GFP11 fragments. These cells constituted the parental cell line used for repeat array insertion.

We then sequentially inserted the loxP-Blasticidin-HSVTK-loxP-TetOx96 or CuOx150-FRT-Neomycin-FRT arrays at each anchor of the targeted loops (**Fig. 1A**). Insertion of the TetOx96 and CuOx150 repair plasmids were selected using 6 µg/mL blasticidin (Invivogen ant-bl-1) and 400 µg/mL G418 (Invivogen ant-gn-5), respectively. The expression of antibiotic resistance genes was designed to direct transcription outwards of the loop interior to avoid interference with loop extrusion. Finally, once TetOx96 and CuOx150 array insertion on the same allele was verified, we removed the antibiotic cassettes using Cre and Flippase (Flp) recombinases. This was done to remove strong transcription of antibiotic resistance genes because it is known that transcription alters chromatin dynamics[30,34] and acts as a mobile barrier of cohesin. 1 µg of Cre and 1 µg of Flippase (FlpO) recombinases were transfected and clones that lost the antibiotic cassettes were selected by the loss of the herpes virus simplex thymidine kinase (HSVTK) gene, making cells sensitive to 8 µg/mL ganciclovir (Invivogen sud-gcv).

Repeat arrays were inserted at the following locations in the human genome: Lv1 (chr8:60,964,180 ; chr8:61,310,370), Lv4 (chr2:235,458,700 ; chr2:236,026,413), Tv1 (chr1:36,980,442 ; chr1:37,901,640), No Loop (chr8:60,733,873 ; chr8:61,310,370), Closed control (chr1:37,900,310 ; chr1:37,901,640).

Auxin-mediated RAD21 degradation and western blotting

To deplete RAD21 fused to the mini auxin inducible degron (mAID)[44], we added auxin (Merck I5148-2G) to a final concentration of 500 µM (from a 500X stock solution diluted in PBS) in fresh culture or imaging medium.

RAD21 depletion kinetics was measured by live-cell imaging and western blotting (**Fig. S3**). For western blotting, cells were grown in 10 cm plates and incubated with auxin for the indicated times. After auxin treatment, cells were washed three times with cold PBS and lysed with 350 µL of HNTG buffer (HEPES pH 7.4 50 mM, NaCl 150 mM, Glycerol 10%, Triton-X-100 1%) with 1X protease inhibitor (Roche 5056489001). After cell collection, lysates were rotated for 30 min at 4°C, sonicated and rotated 30 min at 4°C before centrifugation and supernatants were stored at -80°C until loading. Protein levels were quantified using the Pierce BCA protein assay (Thermo Fisher Scientific, 23225). Samples were boiled for 5 min at 100°C in 1X Laemmli and 10 µg of protein extract were loaded into a 10% Mini-protean TGX gel. Samples were run for 90 min at 110 V and protein transferred to a nitrocellulose membrane for 75 min at 100 V. Membranes were blocked with 5% milk in 1X Tris-Buffered Saline (Tris 20 mM, NaCl 150 mM). Immunostaining was performed with the following antibodies: Rad21 1:1500 (Abcam ab154769), GAPDH 1:50000 (Abcam ab8245), anti-rabbit IR800 1:10000 (7074), anti-mouse IR800 1:10000. We measured fluorescence intensity with the Chemidoc MP Imaging system (Bio-Rad).

For live-cell imaging quantification of auxin-mediated RAD21 degradation kinetics, we used Rad21-mAID-SNAP cells containing the TetR-GFP11x16-NLS and CymR-2xHalo-NLS constructs. Cells were cultured in glass-bottom imaging dishes (Ibidi 81158) for two days. Before imaging, the SNAP JF646 dye was added to fresh medium at a final concentration of 100 nM and cells were incubated for 90 min at 37°C. Cells were washed three times with warm PBS and imaging medium (DMEMgfp (Evrogen MC102) supplemented with 10% FBS or Fluorobrite DMEM (Thermo Fisher Scientific A1896701) supplemented with 1X Glutamax (Thermo Fisher Scientific 35050061) and 10% FBS) was added to the cells. Time-lapse

images were acquired in a bespoke microscope equipped with a 488 nm TA Deepstar Diode Laser (Omicron-Laserage Laserprodukte GmbH) and a 647 nm OBIS LX (Coherent Corp.) for excitation. The microscope was equipped with an Olympus UPLAPO 60x 1.42NA objective, and additional optics leading to a 102 nm pixel size in the final image. Green and far-red fluorescence emission was split at 580 nm by a FF580-FDi02-t3-25×36 dichroic mirror (Semrock) and filtered with 525/50 nm and 685/40 nm fluorescence filters (Alluxa Inc.) respectively. Images were captured by two separate sCMOS cameras: a Zyla 4.2 plus for the green fluorescence and a Zyla 4.2 for the far-red (Oxford Instruments). Sample environment (CO_2 concentration, temperature and humidity) was controlled with a top-stage chamber (Okolab SRL). All devices of the microscope were controlled using python-microscope[45] and using cockpit as graphical interface[46]. We took 31 z-slices separated by 0.4 µm each and z-stacks were taken every 15 min for 45 min at multiple positions. Then, cells were removed from the microscope stage and auxin was added to the medium. Next, we imaged the cells at the same frequency for 4 hours.

To measure RAD21 levels in live-cell images, we used a custom Python script on max intensity projected images. We segmented (with Labkit[47]) and tracked (with TrackMate[48]) nuclei using the green channel containing the TetR-splitGFP-NLS signal. We removed all dividing cells and cells at the edge of the image from the analysis. Using these segmentation masks, we measured the median fluorescence intensity in the RAD21 channel. Background intensity was subtracted, and fluorescence intensities were normalized to the first timepoint of imaging without auxin.

Cell cycle analysis by Fluorescent Activated Cell Sorting (FACS)

To assess the G1 to early S phase fraction of cells, we used FACS to quantify cell cycle phases by propidium iodide staining of fixed cells. Cells were grown in 6-well plates and collected 48 hours later. Cells were trypsinized, centrifuged and resuspended in 500 µL PBS. 4 mL of ice-cold 70% ethanol was added for fixation. Cells were kept at 4°C in 70% ethanol until staining. Cells were washed twice in PBS and incubated for 5 min at room temperature with 50 µg/mL RNAse A solution. Finally, 750 µL of 50 µg/mL propidium iodide solution was added and cells were incubated for 10 min at room temperature.

From the distribution of propidium intensity, the percentage of cells in G1 was computed using the Dean-Jett-Fox model[49] with FlowJo™ software.

Choice of genomic loci for chromatin loop anchor labeling

Using Hi-C, ChIP-Seq and PRO-seq data from [5], we filtered loops and TADs based on the following criteria: size comprised between 300 kb and 1.5 Mb, at least one peak of SMC1, RAD21 and CTCF at both anchors, at least one pair of convergent CTCF sites at anchors, no gene at loop anchors and low gene expression within the loop (<0.5 reads per kilobase million, RPKM) to avoid as much as possible cohesin-independent interactions. In addition, we removed loops that contained co-binding of the three proteins SMC1, RAD21 and CTCF to avoid cohesin pausing during extrusion of the whole loop. From this subset of loops, we removed loops exhibiting enhancers at their anchors (identified in the genehancer double elite set[50]) to ensure that our loops were cohesin-dependent. We then manually removed nested loops and loops containing alignment artifacts in Hi-C maps.

For the closed control cell line, we inserted the TetOx96 repeats 6 kb away (mid-array distance) from the CuOx150 repeat array used to label the 3' anchor of the Tv1 locus. For the no loop control, we inserted the TetOx96 repeats away on the 5' side of the 5' anchor of the Lv1 locus in a cell line that already contained the CuOx150 repeat array at the 3' anchor.

Genomic data analysis

The hg19 genome was used for all genomic analyses. All HCT116 genomic data were retrieved from [5]. Loops and TADs were called using Juicer 1.19.02[51] HiCCUPS and Arrowhead, respectively. The following flags were used for HiCCUPS: -r5000, 0000 -k KR -f 0.1 -p 4,2 -I 7,5 -t 0.02,1.5,1.75,2 -d 20000, 20000; and for Arrowhead: -m 2000 -r 5000 -k KR --threads 10.

For ChIP-Seq data of CTCF, SMC1, RAD21 and NIPBL, we used publicly available data from [5]. Raw reads were quality-checked using FastQC[52]. Reads from different replicates were first mapped independently using bowtie2 v2.2.6.2[53] with default parameters, and the correlation between replicates was computed using wigCorrelate[54]. Replicates with correlations larger than 0.9 were pooled together and mapped again. We remove blacklisted regions[55] and called peaks using default parameters of MACS2 v2.1[56]. CTCF motifs were identified genome-wide using FIMO[57] with the flags -max-stored-scores 50000000 and -thresh 0.001. We then mapped CTCF sites identified with a P-value < $1x10^{-5}$ onto CTCF ChIP-Seq peaks.

ChIP-Seq peaks from CTCF, SMC1, RAD21 and NIPBL were intersected with 20 kb regions centered around the loop anchors using pgltools intersect1D[58].

AB compartments were identified using eigenvector from Juicer with the flags KR BP 100000.

Live-cell imaging of loop anchors

Cells were plated on a 35 mm glass-bottom imaging dish (Fluorodish FD35-100). 48 to 72 hours after seeding, the medium was replaced with fresh medium containing 100 nM of JFX646 Halo dye and the cells were incubated for 15 min at 37°C. Cells were washed twice with PBS and the medium was replaced with live-cell imaging medium (DMEMgfp (Evrogen MC102) supplemented with 10% FBS or Fluorobrite DMEM (Thermo Fisher Scientific A1896701) supplemented with 1X Glutamax and 10% FBS). Before imaging, cells were allowed to equilibrate in the microscopy incubation chamber for at least 15 min at 37°C and 5% CO_2.

Movie acquisition was performed with an inverted microscope (Nikon) coupled to the Dragonfly spinning disk (Andor) using a 100X Plan Apo 1.45 NA oil immersion objective. Excitation sources were a 488 nm (150 mW) and 637 nm (140 mW) lasers. Exposure time was set to 85 ms for both channels with 1.5% laser power in far-red and 5-8% laser power in the GFP channel depending on the imaged cell line. Z-stacks of 29 optical slices separated by 0.3 µm each were acquired every 30 s using the perfect focus system and five different stage positions were imaged for each 2-hour acquisition. The two channels were acquired simultaneously on two distinct EMCCD iXon888 cameras (1024 x 1024 pixels, effective pixel size: 0.121 µm).

Image analysis

Image pre-processing

Images were converted from .ims to .tif format using a custom Fiji macro.

Shift correction

Due to misalignment of the motorized microscope stage device, a small shift can appear between consecutive 3D images. While axial shift can be ignored thanks to the perfect focus device, lateral shifts can reach a few hundreds of nanometers, limiting the efficiency of our tracking analysis. To attenuate lateral shifts, we computed registration parameters between all 3D images from the red channel projected in a 2D space and the first imaged timepoint using 2D cross correlation. To avoid affecting spot detection, we then corrected the displacement on both green and red channels with a pixelic precision to avoid pixel interpolation.

Estimating chromatic aberrations

Correcting chromatic aberrations is crucial to precisely analyze the distance between 2-color loci. To estimate chromatic aberrations in the same plate and media as when imaging the loop anchors, we acquired 3D reference images of actin in WT HCT116 cells in the green and far-red channels using CellMask Deep Red and green actin stains (ThermoFisher A57245 and A57243). At least ten actin images were acquired every two days of imaging. Chromatic shifts were then measured with Chromagnon[59] using the averaged actin images as references and the estimated 3D XYZ translations, 3D magnifications and 2D lateral rotations were used to correct the coordinates of spot localizations throughout the time series.

Eliminating replicated spots

To eliminate replicated spots (*i.e.* spots in cells that, at least, started their S phase), we computed the elongation of the detected spots and manually inspected max-intensity projected movies (**Fig. S2A**). We developed a tool providing a score function of the elongation of the detected spots. A sharp spot might correspond to an isolated spot whereas an elongated spot might correspond to two overlapping spots

coming from replicated DNA. Spot elongation was measured in two steps. First, a 3D Laplacian of gaussian filter was applied on images, and spots were detected by searching for local maxima above a given threshold. Next, we fitted each spot with a 3D second order polynomial function. We measured the amplitude of the paraboloid at 8 positions regularly positioned on a circle around the vertex, and estimated the covariance among these 8 positions. The ratio between the maximum and the minimum of the eigen values minus 1 was used as a score to identify the elongation of the spots. A score closed to 0 corresponded to a sharp spot, whereas a large score corresponds to an elongated spot. We computed this score for each spot and plotted a 2D time- and z-projected map for each movie as a guide to detect replicated spots. By comparing the fraction of cells in G1 (or early S phase) obtained in images and the fraction of G1 cells measured by FACS, we found that our method to eliminate replicated spots was conservative since fractions of cells in G1 were slightly lower in images than in FACS data (**Fig. S2B**).

Tracking fluorescent spots

The 3D image time-series were processed using Fiji. Cells in G1 were encircled by ROIs based on the XYZ elongation of the spots and visual inspection (**Fig. S2**).

Detection and tracking of fluorescent spots were performed separately for each color channel in TrackMate. Spots were detected using the Hessian detector of TrackMate with normalized quality (per ROI) and sub-pixel localization. Detected spots were joined in tracks using the simple LAP tracker from TrackMate with the following parameters: 0.9 µm linking distance, 1.4 µm gap-filling distance and 12 max frame gap. A threshold on the number of spots in each track was then applied to remove tracks emerging from spurious spots. Finally, a second round of detection was performed using the ‘fill gaps’ action in TrackMate with a sphere of radius 0.5 µm,

using interpolated positions as seeds. Detection and tracking parameters were optimized using the Trackmate-Helper plugin on manually annotated images as ground truths.

Pairing of 3D trajectories in the two channels was then performed in Fiji, using the Pair-TrackMate files plugin. Green and red trajectories were paired together by maximizing the number of paired spots under a 2 µm distance threshold. In the cases where several tracks could be paired together, the longest pair in time was chosen.

Trajectory quality filtering

Before analysis, the quality of distance time series was controlled to minimize detection or tracking errors. We filtered track pairing errors by applying a threshold on median track anchor-anchor distance of 0.7 µm in - auxin tracks (except for Tv1 where a 0.9 µm threshold was used), 1.4 µm in + auxin tracks and 0.4 µm in closed control cell line. These thresholds were chosen based on the cumulative distribution function of distances. We filtered out tracks with less than 20 timepoints and removed timepoints where the distance changed by more than 0.5 µm. This removed less than 3% of timepoints. Finally, gaps in tracks were filled by interpolating the distance between the last and next known timepoints. These interpolated distances represented less than 8% of all timepoints.

Measuring localization accuracy

All paired localizations, including interpolated positions, were localized again, but with an optimal method. First, a 1D affine transformation was applied to all pixels to convert the images in photon counts in order to get poisson statistic. The affine transformation was learned by measuring on 100 images of fixed cells the mean and

variance of each pixel, and by fitting an affine curve such as the mean approximates the variance. Then all spots were localized using 3D Gaussian model by Maximum Likelihood Estimation. The spot positions identified by trackmate or by interpolation were used as seeds. This method was used twice. First by fitting the localizations, the amplitudes of the spot, the background and the standard deviation of the model. Second by fitting the localizations and the amplitudes while fixing the standard deviation of the model with the average estimation obtained in the first step. Furthermore, for each localization, we computed the Cramér Rao bound, providing us an estimation of the localization precision in nanometer unit for each spot.

Visualization of images

Solely for visualization purposes, photobleaching of loop anchor movies was corrected by the exponential fitting function of ImageJ.

Imaging data analysis

2-point Mean-Squared Displacement

2-point MSD was computed on the distances as function of time. The exponent was estimated by linear fitting of the logarithm of 2-point MSD in the 30 – 900 s range.

Quantification of closed states

Closed state segmentation

To segment tracks in closed states, we used a simple, model-free approach previously validated on polymer simulations[36]. This method uses a spatial and a temporal threshold. The spatial threshold was defined as a quantile of the closed control distance distribution. To set the temporal threshold, we measured the duration of intervals during which the anchor-anchor distance was always below the spatial threshold (ignoring intervals of 1 timepoint) in RAD21-depleted cells. The

temporal threshold was then defined as a quantile of these time interval distribution. We chose a pair of spatial and temporal thresholds that yielded close to 5% closed state in auxin-treated cells, a temporal threshold lower than 5 min (to avoid missing short closed states as much as possible), an estimated lifetime at least twice higher than the temporal threshold value (to avoid that estimated lifetime is limited by the temporal threshold choice, **Fig. S7F**) and yielded unique results. We used 0.26 µm and 3.5 min for Lv4 and 0.28 µm and 5 min for Tv1 as spatial and temporal thresholds, respectively. Next, we segmented time intervals with anchor-anchor distances below the spatial threshold for durations exceeding the temporal threshold. This allowed us to retrieve a binary trajectory, where 0 indicated no closed state and 1 indicated a closed state. We applied a median rolling filtering to reduce the number of false negative due to brief fluctuations of distances and avoid closed state fragmentation, as validated on polymer simulations[36]. We re-labelled as closed states timepoints exhibiting a value above 0.5. Because of the temporal threshold, our method cannot detect closed states shorter than the temporal threshold.

Computing fraction, frequency and lifetime of closed states

The fraction of closed states was computed as: $\frac{N_{proximal}}{N_{total}}$, where $N_{proximal}$ denotes the number of timepoints spent in the closed state and N_{total} denotes the total number of timepoints (**Fig. 2B**).

The frequency of closed states was computed as: $\frac{N_{occ}}{N_{total}*dt}$, where N_{occ} is the number of occurrences of closed states and *dt* is the frame interval (0.5 min).

The mean lifetime of closed states was computed by exponential fitting of the closed state duration, taking into account censoring as in [17] (**Fig. 2D**).

Bootstraps were computed using 100% of the initial dataset size with replacement.

Estimation of loop state fractions

To estimate loop state fractions, we used a method previously described and validated on simulations (**Figure 3** of Sabaté et al, NAR, 2023)[36] in which an analytical model is fitted to the distribution of coordinate differences. Briefly, the model takes as input the closed and open (auxin-treated cells) coordinate difference distributions, which are known. The extruding distribution is then modelled as an integral over σ^2 varying from σ^2_{closed} to σ^2_{open}. The full analytical model fitted to the data reads:

$$\mathcal{P}_w(R_w;\sigma) = A_{closed}\mathbf{P_w}(R_w;\sigma_{closed}) + A_{open}\mathbf{P_w}(R_w;\sigma_{open}) + A_{extrusion}\int_{\sigma^2_{closed}}^{\sigma^2_{open}}\mathbf{P_w}(R_w;s)ds^2$$

where A_{closed}, $A_{extruding}$, A_{open} are the three state fractions that we estimate, σ^2 is the variance of the coordinate difference distribution, R_w is the anchor-anchor vector and P_w reads:

$$\mathbf{P_w}(R_w;\sigma) = \frac{1}{(2\pi)^{\frac{3}{2}}(\sigma^2+2\sigma_w^2)^{\frac{1}{2}}}exp\left(-\frac{1}{2}\left(\frac{R_w^2}{\sigma^2+2\sigma_w^2}\right)\right), \text{with } w \in \{x,y,z\}$$

To estimate these fractions, the analytical model assumes that extrusion occurs at equilibrium at each step and that the anchors in the extruding state behave as if part of a shorter polymer in which the loop is absent. We used the segmented closed states as input for the closed distribution and the auxin-treated time series for the open distribution.

To compute confidence intervals, we used bootstrapping by randomly drawing time series with replacement from the original dataset of closed, open and untreated time series. The whole set of coordinate differences was used to estimate loop states on

untreated time series, while for closed and open states, the collection of coordinate differences was split in two sets. One set was used to estimate the variance of the coordinate difference distribution for the open and closed states. While the other set was used to infer loop states. This bootstrapping procedure was repeated 1,000 times to compute confidence intervals.

Estimation of DNA loop extrusion speed

To estimate loop extrusion speed, we aligned and averaged segmented closed states on the closed state start and fitted a linear law, as previously described and validated on polymer simulations[36].

$$\langle R^2 \rangle(t) = R_1^2(V_0/s_0)(-t) + R_{\mathrm{loc}}^2$$

Where $\langle R^2 \rangle(t)$ is the mean square anchor-anchor distance, R_1^2 is the mean squared anchor-anchor distance in untreated cells after removal of closed states, V_0 is the extrusion speed, s_0 is the genomic distance between loop anchors and R_{loc}^2 is the localization error contribution. We used R_1^2 as opposed to R_0^2, the mean squared anchor-anchor distance in open states, which was used in the polymer simulation analysis[36]. Since cells were only rarely found in the open state (**Fig. 2E-F**), R_0^2 was rarely reached in the experimental tracks and did not recapitulate the steady-state of loop anchor distance. Therefore, using R_0^2 would lead to an underestimation of the loop extrusion speed.

The loop extrusion speed depends on the window size chosen for fitting[36]. However, using simulations, we showed that the best estimation of the extrusion speed is given by using the window size yielding a plateau in extrusion speed estimations (**Fig. S8**). Thus, our optimal window size was 17.5 min before the closed state start

as this fit window size exhibited a plateau in the estimated extrusion speed depending on the window size (**Fig. S8**).

As negative controls for the extrusion speed estimation, we randomly shuffled distances within single tracks. We then segmented closed states and fitted loop extrusion speed as before.

To compute confidence intervals on the loop extrusion speed, we bootstrapped 1,000 times the closed states, using 100% of the dataset with replacement.

References

1. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
2. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
3. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
4. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6697–E6706 (2018).
5. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
6. Davidson, I. F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822–827 (2023).
7. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).

8. Davidson, I. F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* **22**, 445–464 (2021).
9. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* **15**, 2038–2049 (2016).
10. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
11. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345–1349 (2020).
12. Shin, H. & Kim, Y. Regulation of loop extrusion on the interphase genome. *Critical Reviews in Biochemistry and Molecular Biology* **0**, 1–18 (2023).
13. Zuin, J. *et al.* Nonlinear control of transcription through enhancer–promoter interactions. *Nature* **604**, 571–577 (2022).
14. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
15. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).
16. Xiao, J. Y., Hafner, A. & Boettiger, A. N. How subtle changes in 3D structure can create large changes in transcription. *eLife* **10**, e64320 (2021).
17. Gabriele, M. *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496–501 (2022).
18. Mach, P. *et al.* Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet* **54**, 1907–1918 (2022).
19. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

20. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
21. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641-1659.e26 (2020).
22. Cardozo Gizzi, A. M. *et al.* Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell* **74**, 212-222.e5 (2019).
23. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6**, e25776 (2017).
24. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
25. Mattout, A. & Meshorer, E. Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Curr Opin Cell Biol* **22**, 334–341 (2010).
26. Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* **7**, 540–546 (2006).
27. Sun, L., Fu, X., Ma, G. & Hutchins, A. P. Chromatin and Epigenetic Rearrangements in Embryonic Stem Cell Fate Transitions. *Frontiers in Cell and Developmental Biology* **9**, (2021).
28. Meshorer, E. *et al.* Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. *Developmental Cell* **10**, 105–116 (2006).
29. Grimm, J. B. *et al.* A General Method to Improve Fluorophores Using Deuterated Auxochromes. *JACS Au* **1**, 690–696 (2021).
30. Germier, T. *et al.* Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophysical Journal* **113**, 1383–1394 (2017).

31. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.* **82**, 45–55 (2017).
32. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell* **78**, 539-553.e8 (2020).
33. Rinzema, N. J. *et al.* Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nat Struct Mol Biol* **29**, 563–574 (2022).
34. Shaban, H. A., Barth, R., Recoules, L. & Bystricky, K. Hi-D: nanoscale mapping of nuclear dynamics in single living cells. *Genome Biology* **21**, 95 (2020).
35. Li, S., Prasanna, X., Salo, V. T., Vattulainen, I. & Ikonen, E. An efficient auxin-inducible degron system with low basal degradation in human cells. *Nature Methods* **16**, 866–869 (2019).
36. Sabaté, T., Lelandais, B., Bertrand, E. & Zimmer, C. Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging. *Nucleic Acids Research* **51**, 2614–2632 (2023).
37. da Costa-Nunes, J. A. & Noordermeer, D. TADs: Dynamic structures to create stable regulatory functions. *Current Opinion in Structural Biology* **81**, 102622 (2023).
38. Muniz, L., Nicolas, E. & Trouche, D. RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *EMBO J* **40**, e105740 (2021).
39. Sirinakis, G. *et al.* The RSC chromatin remodelling ATPase translocates DNA with high force and small step size. *EMBO J* **30**, 2364–2372 (2011).
40. Spracklin, G. *et al.* Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. *Nat Struct Mol Biol* **30**, 38–51 (2023).

41. Pradhan, B. *et al.* SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Reports* **41**, (2022).
42. Arnould, C. *et al.* Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660–665 (2021).
43. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research* **47**, W171–W174 (2019).
44. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep* **15**, 210–218 (2016).
45. Susano Pinto, D. M. *et al.* Python-Microscope - a new open-source Python library for the control of microscopes. *J Cell Sci* **134**, jcs258955 (2021).
46. Phillips, M. A. *et al.* Microscope-Cockpit: Python-based bespoke microscopy for bio-medical science. *Wellcome Open Res* **6**, 76 (2021).
47. Arzt, M. *et al.* LABKIT: Labeling and Segmentation Toolkit for Big Image Data. *Frontiers in Computer Science* **4**, (2022).
48. Tinevez, J.-Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
49. Fox, M. H. A model for the computer analysis of synchronous DNA distributions obtained by flow cytometry. *Cytometry* **1**, 71–77 (1980).
50. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, bax028 (2017).
51. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**, 95–98 (2016).
52. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

54. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).

55. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019).

56. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).

57. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

58. Greenwald, W. W. *et al.* Pgltools: a genomic arithmetic tool suite for manipulation of Hi-C peak and other chromatin interaction data. *BMC Bioinformatics* **18**, 207 (2017).

59. Matsuda, A., Schermelleh, L., Hirano, Y., Haraguchi, T. & Hiraoka, Y. Accurate and fiducial-marker-free correction for three-dimensional chromatic shift in biological fluorescence microscopy. *Sci Rep* **8**, 7583 (2018).

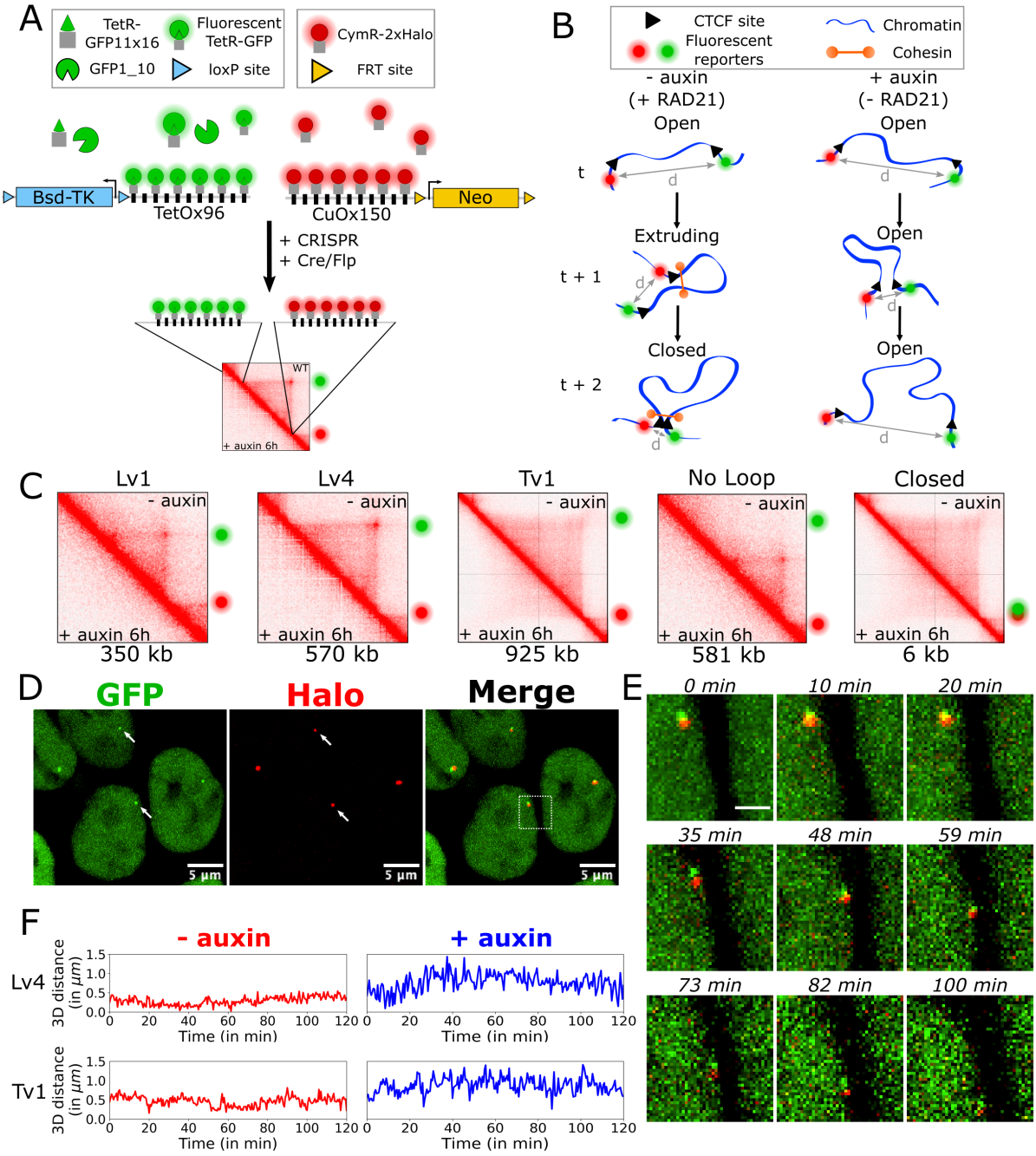


Figure 1: Tracking chromatin loop anchors in living human cells.

A: Loop anchor labeling strategy. TetOx96 and CuOx150 repeat arrays were sequentially inserted in HCT116 cells and visualized using TetR-splitGFPx16 and CymR-2xHalo (JFX646 dye[1]), respectively. Multimerized GFP11 fragments are not shown for clarity. Antibiotic cassettes (Bsd-TK and Neo) were removed by Cre and Flippase (Flp) recombinases. **B:** Cells were imaged with or without auxin, which

depletes RAD21. The distance (d) between the two fluorescent reporters was computed as function of time. In absence of auxin, the chromatin region between the two anchors can be in one of three states: open (no extrusion activity), extruding (*i.e.* containing a DNA loop(s) of growing size(s)), closed (where the two anchors are in close contact). **C:** Hi-C maps of HCT116 cells left untreated or treated with auxin for 6 h. Green and red spots indicate the approximate genomic locations of inserted repeat arrays. The genomic distance between the two fluorescent reporter centers is indicated below each Hi-C map. Hi-C maps are from [2]. **D:** Example images of Lv1 loop anchors at t=0 min. The arrows indicate spots that were not replicated. **E:** Time series of a magnified region corresponding to the dotted white box in **D**, showing the movement of the two Lv1 loop anchors. Scale bar: 2 μm. **F:** Example time series of anchor-anchor distances for Lv4 and Tv1 cell lines in untreated (red) or auxin-treated (blue) cells. All images are maximum intensity projections and were corrected for photobleaching for better visualization.

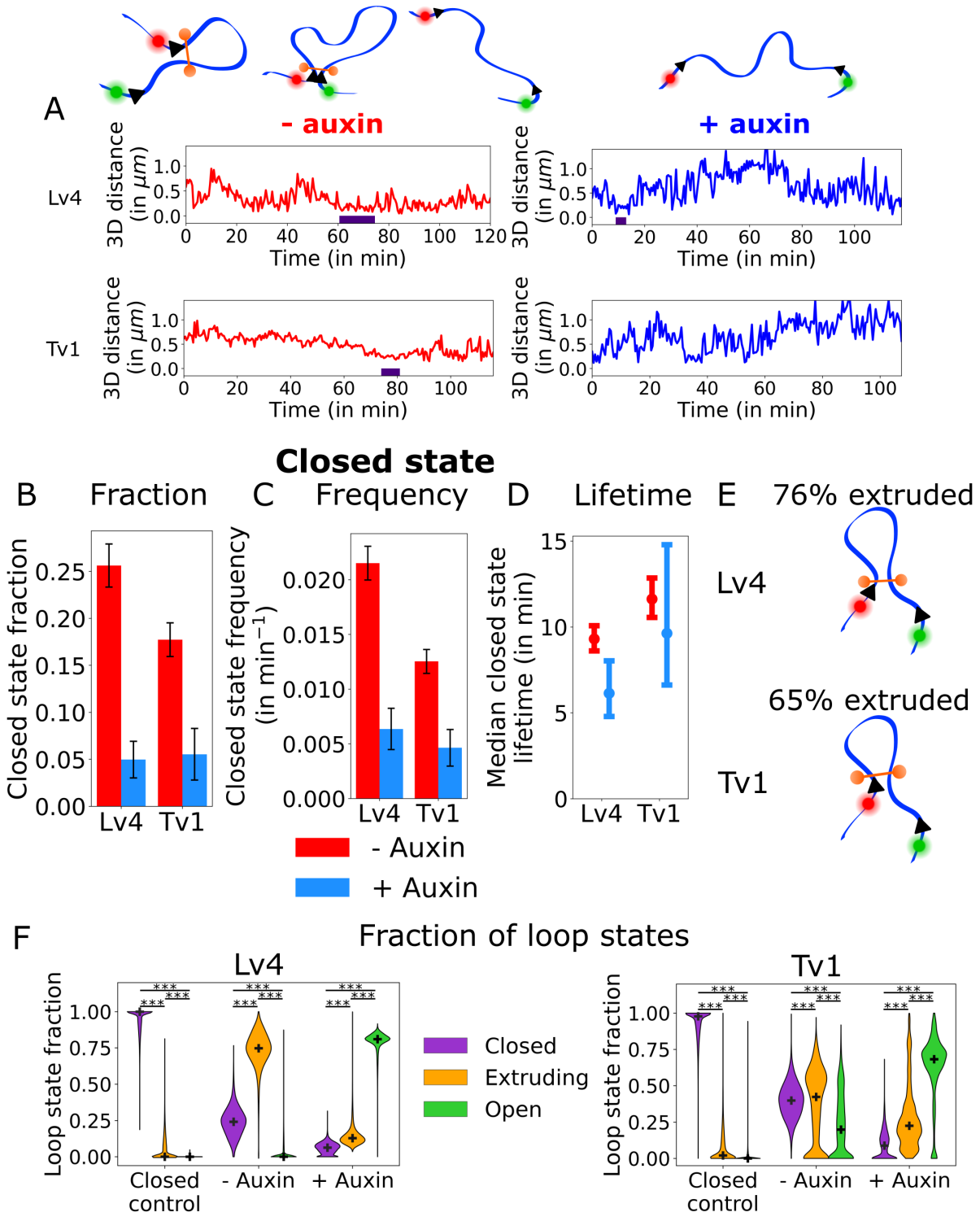


Figure 2: Chromatin loops are dynamic.

A: Example time series of anchor-anchor distances in presence and absence of auxin treatment. The indigo bar indicates time intervals segmented as closed states. **B-D:** Fraction (**B**), frequency (**C**) and lifetime (**D**) of the closed state. Error bars in **B-**

C indicate the 95% confidence interval of 1,000 bootstrapped samples. Error bars in **D** indicate the 95% confidence interval of the exponential fit used to compute closed state lifetime. Red and blue colors indicate cells without or with auxin treatment, respectively. **E:** Fraction of TAD extruded computed as 1 - the ratio of mean squared distances in untreated and auxin-treated cells **F:** Fraction of loop states for Lv4 (left) and Tv1 (right) cell lines. For each cell line, the fraction of loop states in the closed control, untreated and auxin-treated cells is shown. Purple, orange and green colors indicate the fraction of closed, extruding and open states, respectively. ***: P-value < 0.001 from the post hoc Dunn's test performed after Kruskal-Wallis test. N=50,000 bootstrapped samples

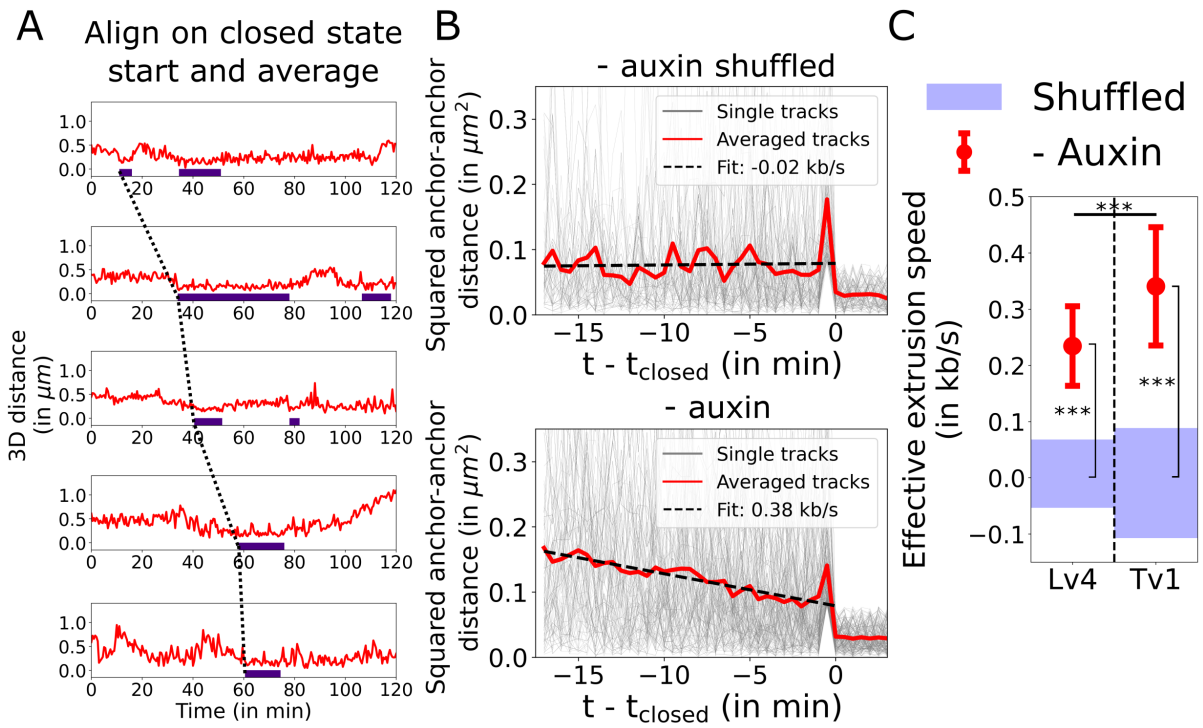


Figure 3: Chromatin loops are effectively extruded at 0.3 kb/s in living human cells.

A: Time series of 3D anchor-anchor distances in untreated cells. Indigo bars indicate segmented closed state intervals. Dotted lines indicate how the starting times t_{closed} of segmented closed states are aligned across single time series. **B:** Time series of squared anchor-anchor distances obtained after alignment of closed state starting times, with (top) or without (bottom) random shuffling of time points. N=319 time series (grey) are averaged (red) to linearly fit (black dashed line) the effective loop extrusion speed. **C:** Estimated effective loop extrusion speed. Blue shaded area represents the 95% confidence interval of the estimated speed in randomly shuffled time series. Error bars are 95% confidence interval of n=1,000 bootstrapped samples. ***: P-value<0.001 for a Tukey's test preceded by an ANOVA. *** between error bars and blue shaded area indicate significant differences between extrusion speeds estimated from untreated and shuffled time series.

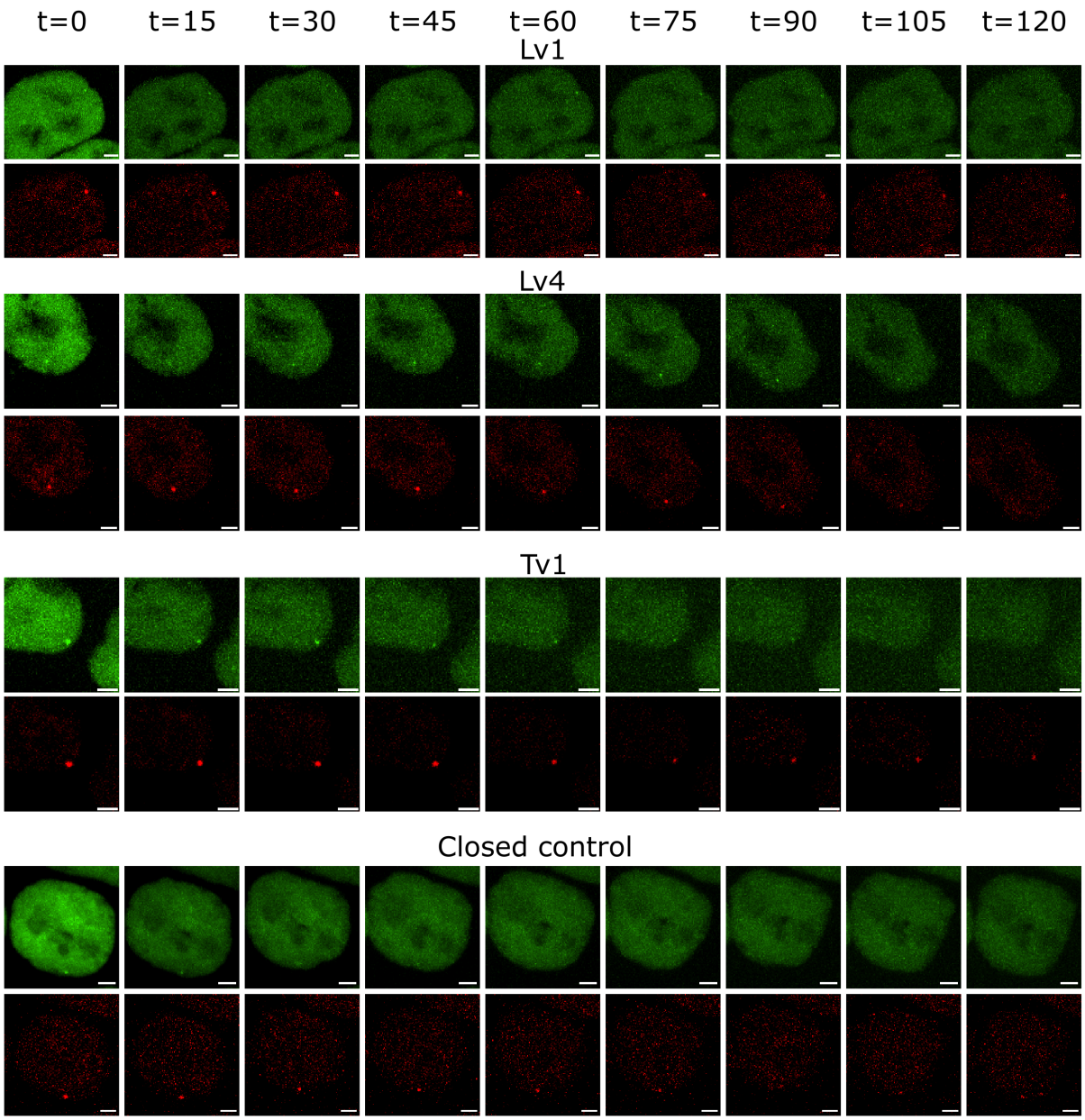


Figure S1: Example time-lapse images of each cell line.

Examples of time-lapse imaging data for the four cell lines Lv1, Lv4, Tv1 and the closed control in the two color channels. Image intensities were corrected for photobleaching using an exponential fit for visualization purposes. Maximum intensity projections are shown. Timepoints are indicated in minutes. Scale bar: 2 μm.

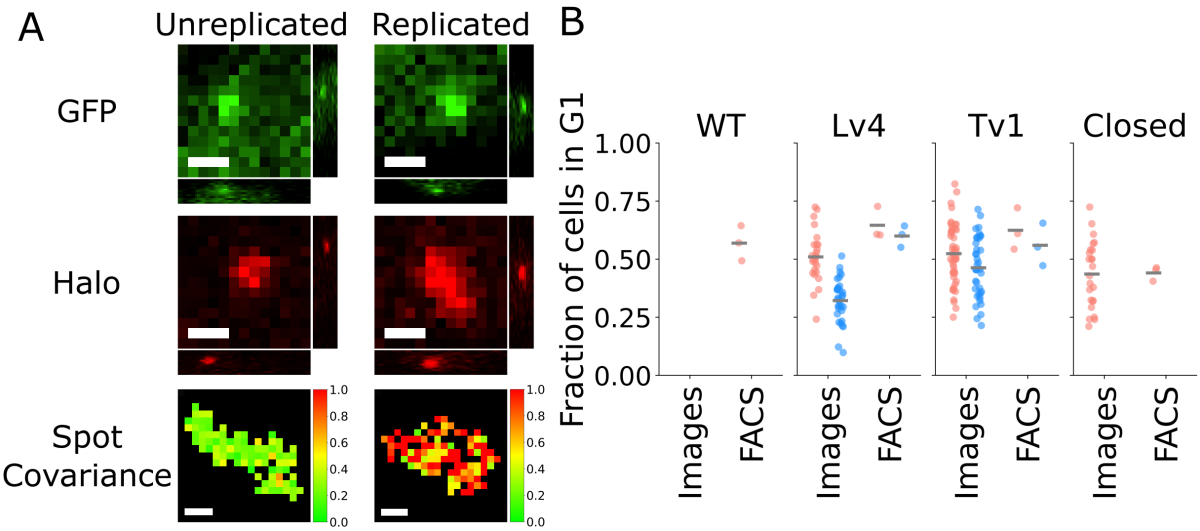


Figure S2: Elimination of cells in S or G2 phases.

A: Example images of unreplicated (left) and replicated (right) fluorescent spots in the two fluorescent channels. A single z-plane centered on the spot center is shown. The spot elongation was computed as a measure of spot spreading and projected in z and time (bottom). Scale bar: 0.5 μm. **B:** Fraction of cells in G1 or early S phase as determined by imaging or FACS analysis. For images, the fraction of cells remaining after elimination of cells exhibiting replicated spots is shown. Each dot shows the fraction of cells in G1 (or early S phase) for one field of view. For FACS data, each dot shows the G1 fraction of one replicate. N=3 replicates for FACS data.

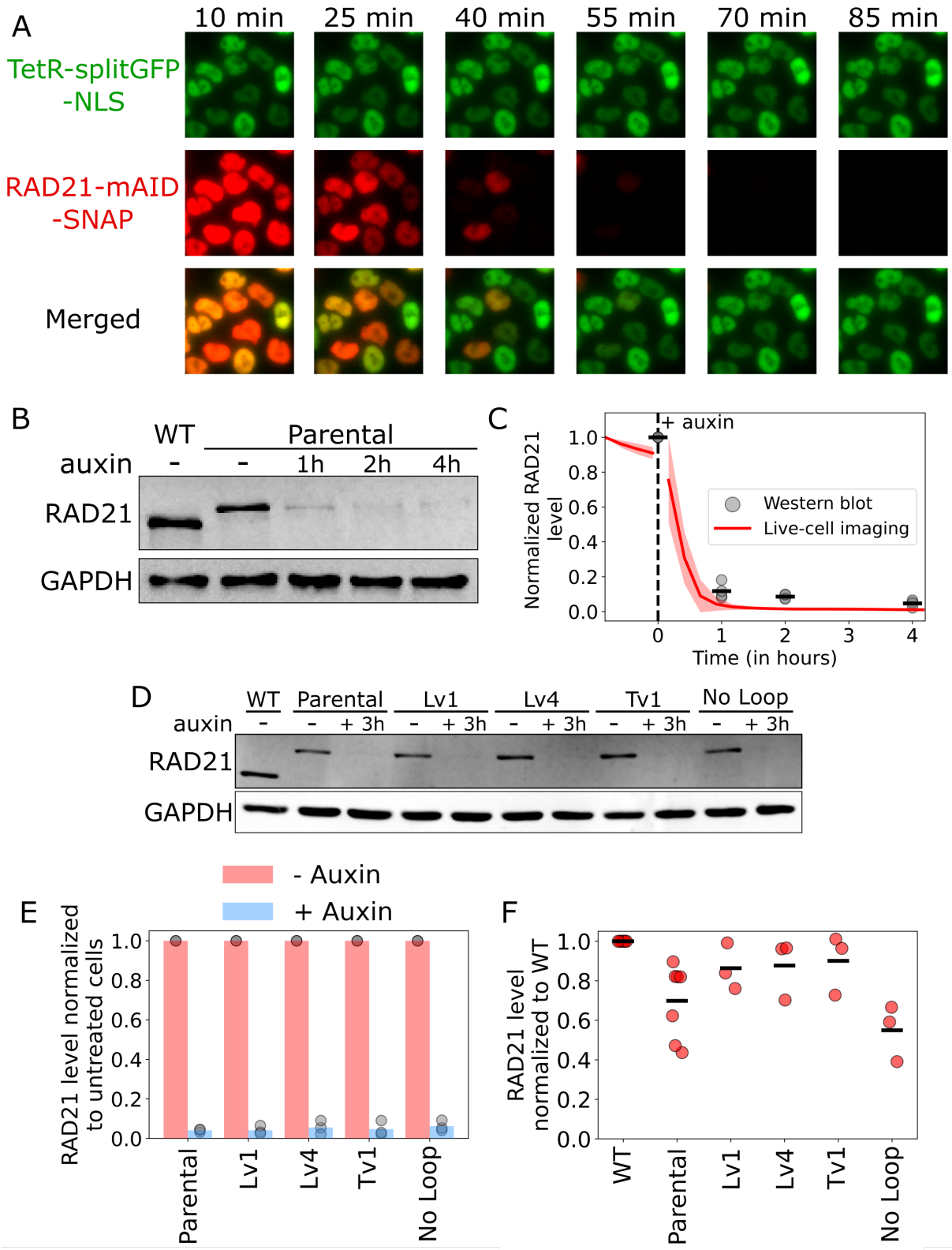


Figure S3: RAD21 degradation kinetics using the auxin-dependent AtAFB2 degron.

A: Live cell images of cell nuclei after auxin treatment. TetR-split-GFPx16-NLS (green) was used to segment nuclei. Time after auxin addition is indicated on top. **B:**

Western blot of the kinetics of RAD21 depletion upon auxin treatment in parental cell lines. **C:** Quantification of RAD21 level as function of auxin treatment duration from live-cell imaging (red line) and western blot (grey dots). For live-cell imaging, the RAD21-mAID-SNAP median fluorescence intensity inside segmented nuclei is shown. Each timepoint shows the median RAD21 intensity over all cells from a single replicate. The red shaded area shows the 95% confidence interval. N=4 replicates (165-596 cells per replicate). For western blot, RAD21 level was normalized to GAPDH. Horizontal black lines indicate the mean of N=4 replicates. The black dotted line shows the timepoint at which auxin was added to the medium. **D:** Same as **C** for each cell line containing repeat arrays with or without a 3-hour auxin treatment. **E:** Western blot quantification of RAD21 degradation in the different cell lines without (red) or with (blue) a 3-hour auxin treatment. Bars indicate the mean of N=3 replicates. **F:** Quantification of basal degradation in untreated cell lines as compared to the WT RAD21 level. Black horizontal lines indicate the mean of N=3 replicates (N=7 replicates for the WT and parental cell lines).

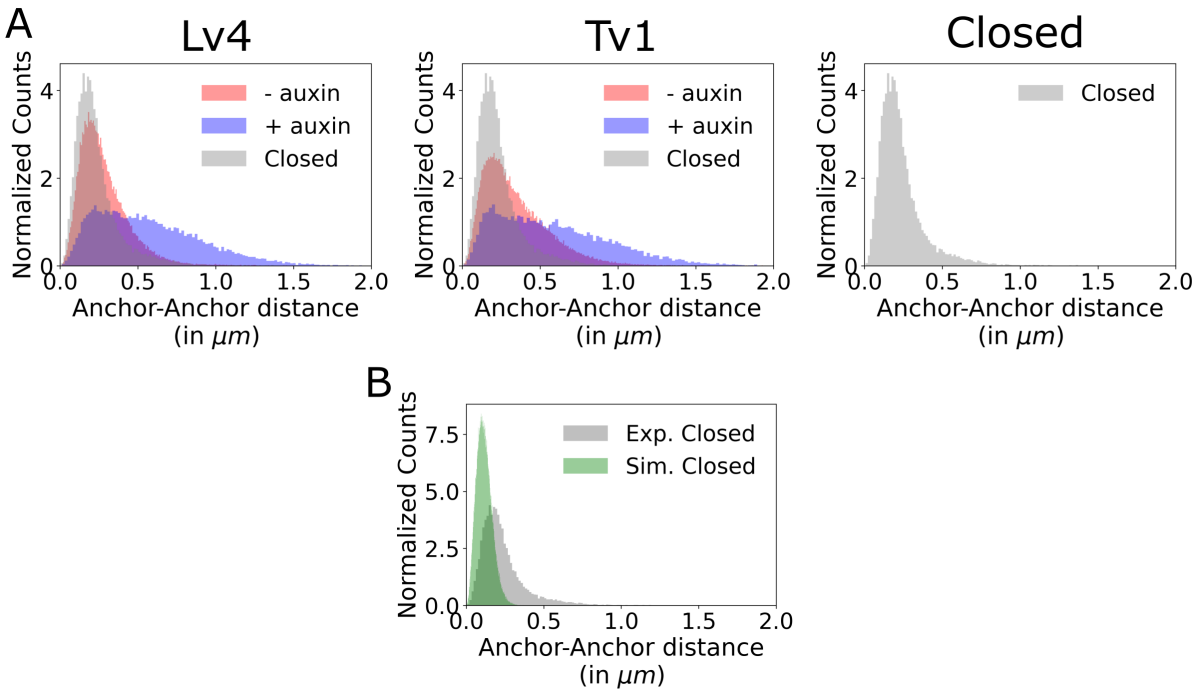


Figure S4: Distributions of anchor-anchor distances, with and without auxin.

A: Histograms of anchor-anchor distances for the cell lines Lv4, Tv1, with (blue) or without (red) auxin treatment and the closed control, with reporters separated by only 6 kb (grey). The histogram of the closed control is shown superposed in grey for all cell lines. **B:** Green histogram shows anchor-anchor distances predicted by polymer simulation assuming 6 kb between reporters (as in the closed control experimental cell line) and the maximal localization errors measured in this cell line. Despite these inflated localization errors in the simulation, the simulated distances are much smaller than the experimental ones (grey) for a reason yet to be determined.

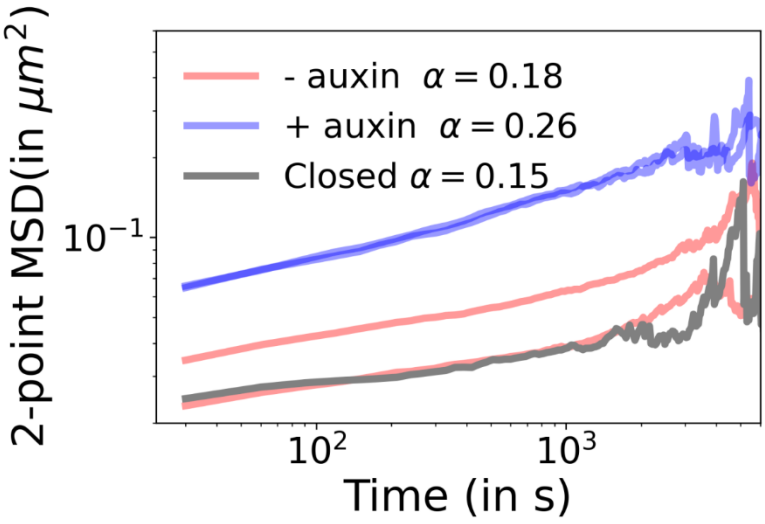


Figure S5: RAD21 decreases chromatin motion.

2-point Mean Squared Displacements (MSD) for untreated (red), auxin-treated (blue) Lv4 and Tv1 cell lines, and the closed control (grey) cell lines are shown. The untreated cells exhibited the α=0.2 exponent observed previously[3], while this exponent increased after RAD21 depletion.

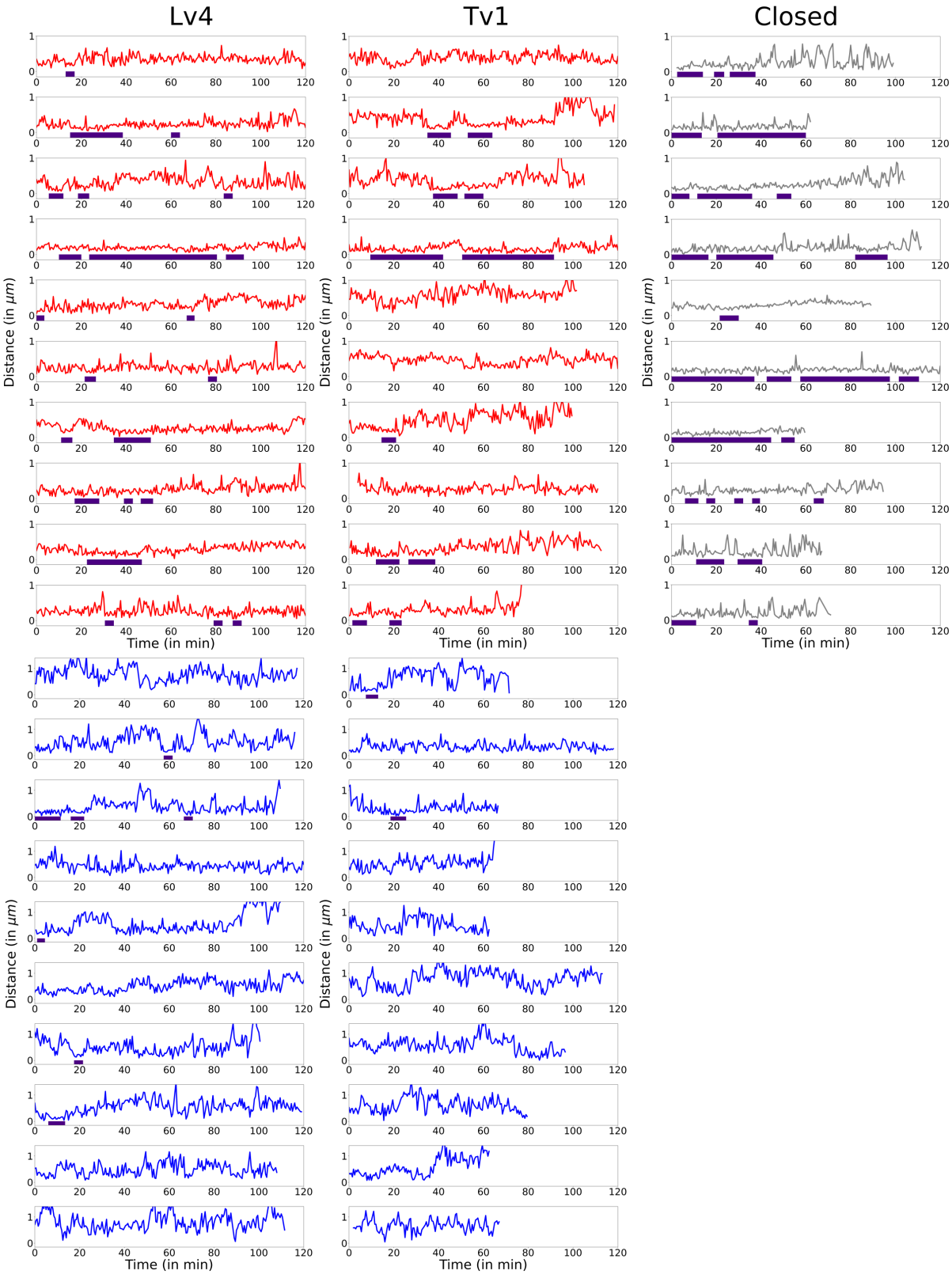


Figure S6: Example time series of anchor-anchor distances and closed state segmentation.

Red time-series show untreated cells, blue time series show auxin-treated cells. The untreated closed control is shown in grey. Segmented closed state intervals are indicated by indigo bars. Auxin-treated cells were plotted with a vertical axis limit of 1.4 μm, instead of 1 μm for other tracks.

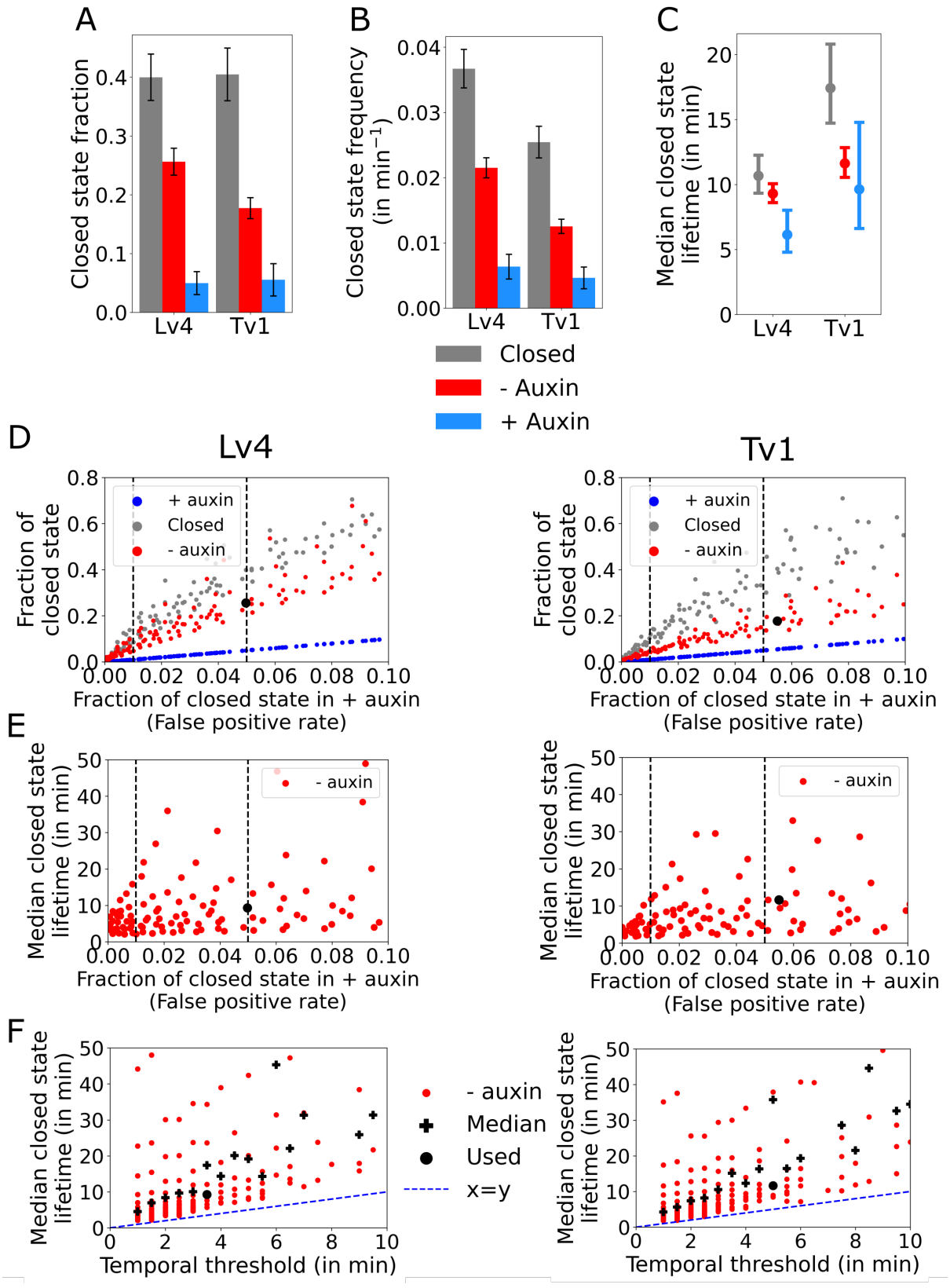


Figure S7: Quantification of closed loop states for different pairs of spatial and temporal thresholds.

A-C: Fraction (**A**), frequency (**B**) and lifetime (**C**) of closed states for Lv4 and Tv1, as in **Fig. 2B**, but including the closed control. Error bars in **A** and **B** indicate 95% confidence interval of n=1,000 bootstrapped samples, while they indicate the 95% confidence interval of the exponential fit in **C**. **D-E:** Fraction of closed states (**D**) or closed state lifetimes (**E**), for various pairs of spatial and temporal thresholds used in closed state interval segmentation, as function of the assumed fraction of falsely detected closed states in the auxin-treated cells (false positives). Black dotted lines correspond to 1% and 5% false positives. **F:** Median closed state lifetimes estimated using various pairs of spatial and temporal thresholds, as function of the temporal threshold. For each temporal threshold, black crosses indicate the median lifetime over all spatial thresholds. In **D-F**, each dot corresponds to a pair of spatial and temporal thresholds and the black dot indicates the pair of thresholds used to segment closed states in **Fig.2** and **Fig. 3**. In **E-F**, the vertical axis was cut at 50 min for legibility.

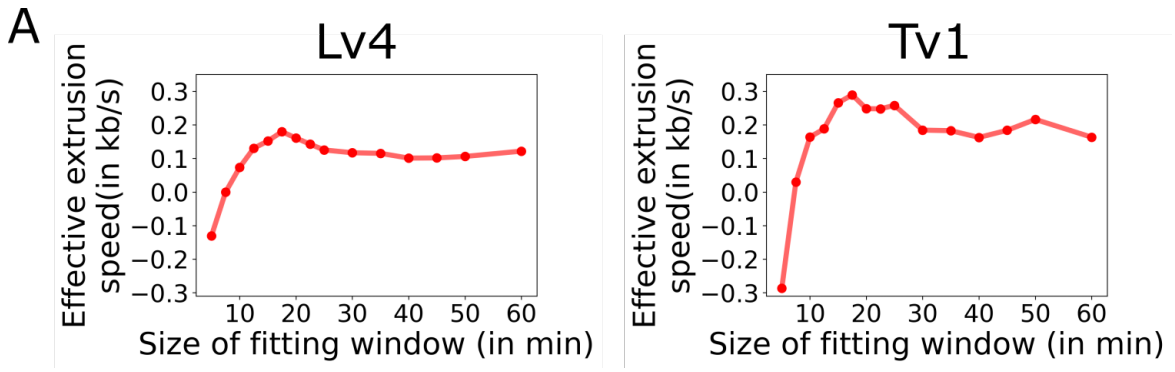


Figure S8: Effective loop extrusion speed as function of the size of fitting window before closed state start.

A: Estimated effective loop extrusion speed as function of the fit window size. Loop extrusion speed reaches a maximum depending on the fit window size, as predicted by simulations[4], using the timepoints up to 17.5 min before the closed state. No bootstrapping was performed to estimate loop extrusion speed.

Condition	Number of movies	Number of single time series	Average number of frames per time series (without interpolation)	Average length of time series (in min)	Average random localization errors (σ_X, σ_Y, σ_Z) in nm (green channel)	Average random localization errors (σ_X, σ_Y, σ_Z) in nm (far-red channel)	Error on 3D distance (in nm)
Lv4 -auxin	41	427	147	73	27, 27, 57	13, 13, 28	77
Lv4 +auxin	13	154	124	62	NA	NA	NA
Tv1 -auxin	45	606	118	59	32, 32, 75	12, 12, 27	94
Tv1 +auxin	16	126	89	44	NA	NA	NA
Closed -auxin	31	168	90	45	38, 38, 82	14, 14, 29	104

Table S1: Overview of imaging experiments and analyses.

All statistics are from quality filtered datasets. NA: Not assessed. The error on distance was computed as: $\sqrt{\sigma_{x,green}^2 + \sigma_{y,green}^2 + \sigma_{z,green}^2 + \sigma_{x,far-red}^2 + \sigma_{y,far-red}^2 + \sigma_{z,far-red}^2}$.

References:

1. Grimm, J. B. *et al.* A General Method to Improve Fluorophores Using Deuterated Auxochromes. *JACS Au* **1**, 690–696 (2021).
2. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
3. Mach, P. *et al.* Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet* **54**, 1907–1918 (2022).
4. Sabaté, T., Lelandais, B., Bertrand, E. & Zimmer, C. Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging. *Nucleic Acids Research* **51**, 2614–2632 (2023).

3.2 Short-term perspectives and limitations on chromatin loop dynamics quantification

3.2.1 Short-term perspectives

Characterization of the unexpected behavior of the closed control cells and expectations on the analysis of the no loop cell line

Below, I detail possible explanations and future work aiming at understanding their behavior.

First, our closed state control where fluorescent reporters are 6 kb apart exhibited an unexpectedly large distribution of distances. The distribution of distances was much larger than for a simulated polymer, even after adding the highest measured experimental noise to the simulated distances (**Fig. S5B** of Sabaté *et al*, unpublished). A similar issue was raised by two previous studies that used the same type of control[163,164]. A possible explanation is chromatin decondensation at the location of repeats. Indeed, within each array, repeats are separated by less than 14 bp and bound by their respective protein (TetR or CymR). If all repeats are bound at the same time, nucleosomes may not bind DNA, which is therefore decondensed and could exhibit relatively large spatial distances, even for short genomic distances (6 kb here). However, this cell line could not be fully genotyped and might have undergone unexpected genomic rearrangements. I plan to analyze a second cell clone of the closed control to check whether this behavior is reproducible.

Second, the 'No Loop' control cell line, where one of the repeats is not inserted at a loop anchor, but rather at a genomic locus without a specific feature in Hi-C maps (**Fig. 1C** of Sabaté *et al*, unpublished), should greatly help to choose the thresholds used for closed state quantification and help validate further the analysis method. This cell line represents a very valuable control. Since no closed state is expected in this cell line, we could use it to set the false positive rate in auxin-treated cells used to segment closed states.

Capture Micro-C maps of the studied loci

The Hi-C maps shown in the above study are from [31]. These maps were built using HCT116 cells with a RAD21 degron but not the exact same construction as I used. We plan to perform a Capture Micro-C protocol to reveal the topology of our examined loci in untreated and auxin-treated cells, in collaboration with Giacomo Cavalli's lab. This should allow us to compare Hi-C patterns with the estimated dynamics of each locus.

Increasing localization precision and robustness of track quality-filtering

Before analysis, tracks were filtered based on several criteria to remove as much as possible tracks and timepoints emanating from spurious detections. I plan to improve the localization precision by gaussian fitting of fluorescent spots. In addition, this would allow me to compute a localization precision for each fluorescent spot, at each timepoint. This precision quantification should enable a less-biased filtering method and will allow to take into account localization precision in downstream analyses. For instance, timepoints with low localization precision could be assigned a lower weight in the loop extrusion speed fitting procedure.

How can differences in loop dynamics between genomic loci be explained?

With only two different loci analyzed, I refrained from drawing conclusions about the influence of parameters such as genomic distance or Hi-C peak intensity on loop dynamics. However, when ready, the addition of the Lv1 and 'No Loop' cell lines will enable a deeper and more refined analysis of our live-cell imaging data. First, I plan to use modeling to infer parameters of loop extrusion such as cohesin processivity and density, and the permeability of CTCF barriers. By comparing modeling and experimental estimates of loop dynamics, and by combining the results from the different loci to constrain the parameter space, I expect to reveal whether or not the same extrusion parameters can explain loop dynamics observed at different genomic loci. Second, the different loops quantified here exhibited different fractions of loop states as well as different frequencies and lifetimes of closed states. It is not yet known what genomic features determine loop dynamics. Using the above-mentioned analysis, 1D genomic features (*e.g.* CTCF site strength and location and ChIP-Seq data of extrusion-related proteins), Hi-C contact maps and live-cell quantifications, I aim to define the genomic determinants of loop dynamics. Depending on the variability of loop dynamics, a few to numerous loops quantified in living cells would be needed for a thorough quantification, but using the four available cell lines should provide insights into the role (or absence of role) of genomic features in loop dynamics.

A more refined analysis of loop extrusion speed yields similar results

In our simulation analysis, we noticed that adding localization errors to simulated anchor-anchor distances tended to infer the closed state starts earlier than the correct starting timepoints. For this reason, we developed an algorithm that can re-estimate the start of the

closed state by applying small temporal shifts to the time series relative to their closed state segmentation. This allowed us to locate closed states that would start between two acquired frames and correct small errors in closed state segmentation due to the finite localization precision. This more complex analysis and the simpler analysis presented here resulted in similar extrusion speed estimates. I consequently only presented the simplest analysis.

To estimate loop extrusion speed, we assumed that the mean squared anchor-anchor distance decreases linearly before reaching the closed state. This was verified in simulations where a single cohesin molecule could switch from bidirectional to unidirectional extrusion upon reaching an anchor, therefore exhibiting two different speeds[101]. However, we plan to verify that this progressive decrease of mean squared anchor-anchor distances remains approximately linear over a much wider range of cohesin density and processivity.

3.2.2 Limitations

RAD21 depletion provides a pure open state control but lacks loop extrusion

The quantification of loop extrusion dynamics requires the use of different controls. I chose to deplete RAD21 to eliminate all loops from the genome and measure distances in a purely open state. However, a CTCF depletion, or more accurately, a mutation of CTCF sites at loop anchors would help for some analyses. Indeed, CTCF depletion leads to the loss of the closed state, but unlike RAD21 depletion, allows extrusion to proceed, thereby allowing for both open and extruding states. Therefore, using a CTCF instead of a RAD21 depletion would increase the precision of closed state segmentation by removing only the closed state and not both the closed and extruding states, as with RAD21 depletion. However, since CTCF depletion contains a mix of open and extruding states, the purely open distribution would not be known, while it is required and very informative in several analyses that we performed. Building cell lines where both proteins can be depleted would be the ideal case (*e.g.* using two orthogonal degrons[303]). However, making such cell lines and analyzing multiple labelled loops would require considerable additional efforts.

What type of closed state control should we use?

In addition to the purely open state control, we needed a purely closed state control. Ideally, this control would consist of a cohesin-mediated loop blocked in its closed state. However, it is unclear how to obtain this experimentally yet, therefore I had to choose a

proxy. I inserted the two fluorescent reporters 6 kb apart (mid-array distance) as a proxy for the closed state. The 6 kb takes into account the genomic separation between loop anchors and reporters present in the analyzed cell lines. In the cohesin-mediated closed state, the expected distance between the anchors is likely not zero but potentially up to ~50 nm, which is the size of the cohesin ring. Other types of closed controls could be considered. A WAPL depletion should increase the closed state fraction but is not expected to yield a pure closed state population. A control where cohesin becomes covalently linked to CTCF upon their encounter could be interesting. However, such a system has yet to be developed. Alternatively, closed state controls could be replaced by theoretical predictions such as in [164].

Inferring dynamic parameters from the comparison of polymer simulations and experiments

Previous studies aiming at quantifying loop extrusion dynamics in living cells inferred multiple parameters based on the comparison of polymer simulations and experimental results (**Figure 20**)[163,164]. I also plan to use this approach to estimate indirectly hidden parameters of loop extrusion such as cohesin processivity and density, CTCF permeability, *etc*. However, estimates inferred from such comparisons need to be examined with caution. Indeed, models assume several hypotheses that may be violated in experimental conditions. For example, chromatin might not reach equilibrium during loop extrusion. Therefore, we currently work on several new analyses to extract as much information as possible from the experimental data alone.

Cancer cell lines do not reflect the behavior of primary cells

I used HCT116 cells, a near-diploid colorectal cancer cell line. Cancer cells are convenient to work with due to their immortality and high doubling rates. However, cancer cells are aberrant cells that can exhibit pluripotency features[304,305] and differ from primary cells in their genome architecture[306,307]. Results derived from such cell lines should thus be taken with caution. Therefore, loop extrusion dynamics might differ in primary cells as compared to our estimates in HCT116 cancer cells.

3.3 Identifying genomic determinants of loop dynamics by high-throughput DNA FISH

Despite several genomic studies investigating the interplay between loop extrusion and other genomic features, the influence of chromatin state on loop extrusion remains poorly

characterized[31,61,76]. Although live-cell imaging gives access to a rich information, only a few genomic loci can be quantified due to the long and tedious process of CRISPR-based genome editing. Since each genomic locus harbors distinct and unique chromatin states, the analysis of only a few genomic loci might not allow generalizations to the whole genome. To identify the genomic determinants of loop extrusion dynamics, I tried to quantify the fractions of open, extruding and closed states using high-throughput DNA FISH of loop anchors on hundreds of genomic loci. I wished to image loop anchors located in different chromatin states, displaying different loop strengths and insulation, as well as different levels of gene expression and CTCF binding. Using the static analysis developed on chromatin simulations (**Fig. 4** of Sabaté *et al*, NAR, 2023)[101], this should in principle allow us to define the variability of loop structures as function of chromatin states.

For this purpose, I decided to set up a high-throughput DNA FISH protocol to image loop anchors in hundreds of genomic loci (**Figure 18**). I thoroughly tested the different steps of DNA FISH protocols (permeabilization; alkaline, heat denaturation or single-strand exonuclease resection[308]; duration of denaturation; salt concentration during washes; number of probes and length of targeted genomic regions), but I could not obtain a satisfying protocol to work. I designed probes (200 probes / locus) against 25 kb regions located at loop anchors for dozens of different genomic loci. Despite improvements in this final design as compared to previous ones, the resulting image quality was highly variable across different loci (**Figure 18B**). The fluorescent background was often high, limiting our ability to detect spots. I expected to observe 2 loci per nucleus (or 2 spatially close pairs of spots after replication), but I could observe nuclei with more than 2 spots (~31% of targeted loci displayed approximately 2 spots per nucleus) and some spots in one channel were not spatially close to spots in the other channel, suggesting unspecific labeling or other artefacts (**Figure 18B**).

Each locus in the genome harbors specific chromatin features and thousands of loops are identified genome wide. While a high-throughput live-cell imaging approach is technically difficult to perform and requires more work, high-throughput imaging of loop anchors by DNA FISH should give sufficient information to characterize loop structures in different chromatin contexts.

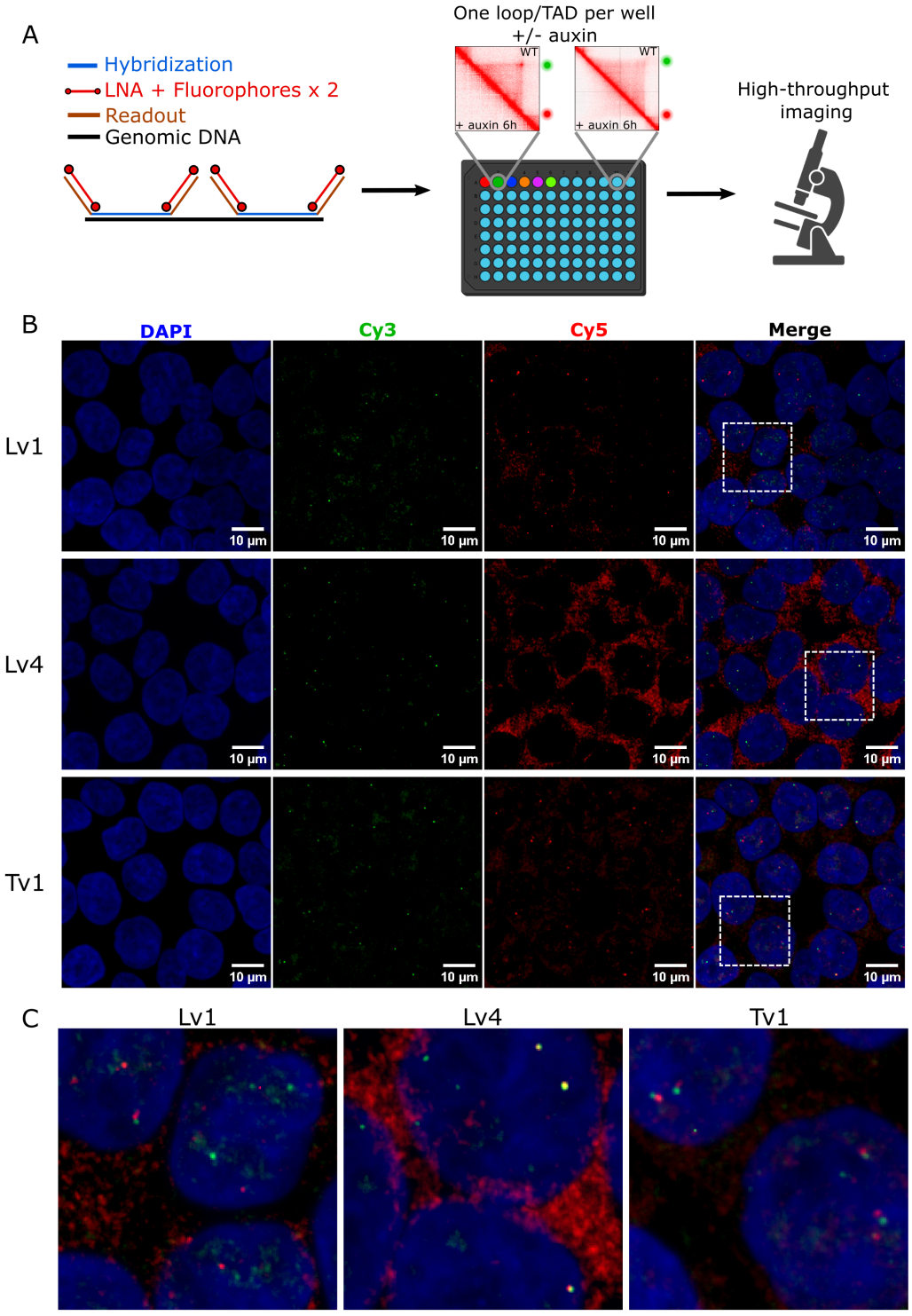


Figure 18: An approach for high-throughput DNA FISH on hundreds of chromatin loops and TADs. A: Experimental design. Probes are amplified from an oligopool and hybridized on DNA. Their location is revealed by fluorescent LNA bound to reader sequences. Cells are treated or not with auxin to deplete cohesin. DNA FISH is performed in 96-well plates, where each well contains the probe to image both loop anchors of a unique locus. Cells are imaged with a high-throughput microscope. **B:** Representative DNA FISH images of 3 different loops (Lv1, Lv4 and Tv1 imaged in living cells). Two fluorescent spots per nucleus are expected. **C:** Zoomed images from the white dotted rectangles in **B**.

Discussion

1. Summary

Chromatin organization and the process of chromatin loop extrusion is being increasingly characterized by a large panel of techniques. This process is tightly regulated at several scales. However, despite our knowledge on its regulation, we lack information about the dynamics of chromatin loops. To address this gap, we decided to track multiple endogenous chromatin loop anchors in living human cells and quantify their dynamics.

Because of stochastic motion of chromatin and unavoidable errors in computing fluorescent spot localization, it is not *a priori* evident that loop extrusion can be detected and quantified by tracking loop anchors. We used polymer simulations to determine the experimental conditions under which one can expect to detect and quantify loop extrusion dynamics from measuring the distance between loop anchors. Furthermore, we tested and validated novel analysis methods to quantify loop dynamics such as: the fraction of closed, extruding and open states from static imaging only, the fraction and lifetime of closed states and the effective loop extrusion speed *in vivo* from dynamic imaging.

Having established that chromatin loop dynamics can be quantified under realistic experimental conditions, we labelled several pairs of endogenous loop anchors and measured their distance as function of time in living HCT116 cells. We found that loop anchors occasionally contacted each other (7-27% of closed state fraction). These contacts were transient (2-16 min) as compared to the 10-hour G1 phase duration and occurred 0.8-1.3 times per hour). While contacts between loop anchors were short-lived, we found that loops were generally found in an actively extruding state and never in a purely open state. Finally, we estimated the effective *in vivo* loop extrusion speed to be ~0.3 kb/s.

These results suggest that cohesin-mediated loops are dynamic during the G1 phase of the cell cycle and that their anchors only contact each other transiently despite loops being almost constantly extruded (**Figure 19**). In addition, the continuous extrusion of loops creates frequent contacts within the loop, besides loop anchors, which possibly favors functional interactions between DNA elements located within the loop. Thus, the extrusion process transiently increases long-range chromatin interactions rather than creating prolonged and

stable contacts. The dynamic process of loop extrusion might be more essential to their biological function than the loops themselves.

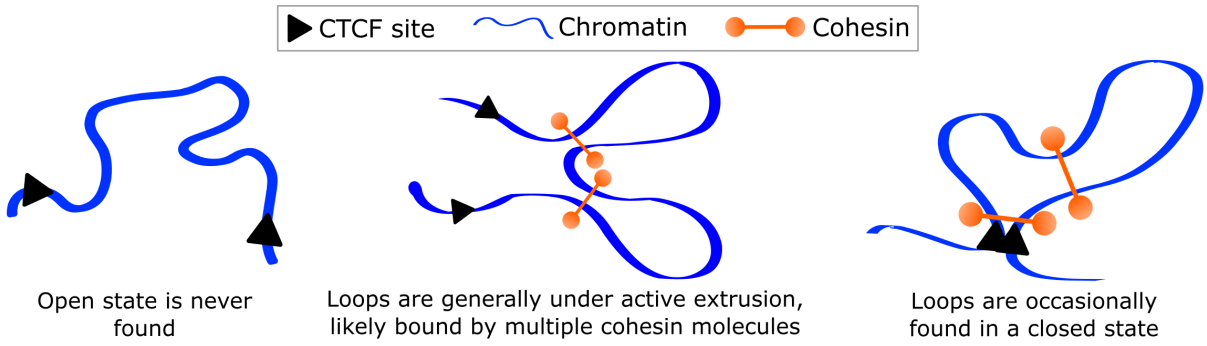


Figure 19 : Chromatin loops are dynamic structures. We found that cohesin-mediated loop extrusion creates loops that are never devoid of cohesin molecules. Loops are instead generally under active extrusion, likely bound by several cohesin molecules simultaneously. Loops were occasionally found in a closed state. **Figure 20** summarizes our numerical estimates of chromatin loop dynamics.

2. Chromatin loops are consistently dynamic across model systems

2.1 Chromatin loops are dynamic structures, constantly subjected to extrusion

During the course of my PhD, two different studies aiming at quantifying loop extrusion dynamics by tracking loop anchors in living cells were published[163,164]. Below, I describe, compare and discuss the results of their and my efforts to quantitatively characterize the loop extrusion process. A synthetic comparison is provided in **Figure 20**.

2.1.1 Experimental design for the quantification of loop extrusion

The experimental design is similar between the three studies: fluorescent reporters were inserted at loop anchors and were tracked as function of time in living cells (mESCs in Mach *et al* and Gabriele *et al*, HCT116 human cancer cells in our study). The temporal resolution at which loop anchors were tracked was also similar between the studies: one 3D image per 30 s for 3 hours (Mach *et al*), one 3D image per 30 s for 2 hours (my study); one 3D image per 20 s for 2 hours (Gabriele *et al*) (**Figure 20**). While Gabriele *et al* and I inserted fluorescent reporters at endogenous genomic loci, Mach *et al* used a synthetic strong triplet of CTCF sites inserted in a TAD where internal CTCF sites were previously deleted. In addition,

Mach *et al* placed the fluorescent reporters within the loop (where repeat arrays could potentially interfere with loop extrusion), whereas Gabriele *et al* and I placed them outside the loop (**Figure 20**). In all studies, labelled loops were devoid of gene expression, to avoid non-cohesin related interactions and quantify purely extrusion-related motion. The genomic size and strength of the 'dot' interaction at TAD corner varied between studies. Mach *et al* and Gabriele *et al* both visualized a single locus: a small 150 kb loop for Mach *et al*, and a 505 kb loop for Gabriele *et al*. In this work, I quantified two genomic loci: a 570 kb loop with a strong dot at the TAD corner and a 925 kb weak TAD (**Figure 20**). I also generated cell lines to visualize two additional genomic loci, which we plan to visualize and quantify in the near future. Finally, in our study we depleted solely RAD21, while Gabriele *et al* depleted RAD21, CTCF and WAPL and Mach *et al* used RAD21 depletion and deletion of CTCF sites. Although Gabriele *et al* and Mach *et al* had access to a richer information than us thanks to their different depletions, they could only quantify a single locus, while we were able to quantify loop dynamics on multiple genomic loci with the aim of comparing the variability of their dynamics.

Gabriele *et al* and our study used a proxy for closed states. Gabriele *et al* and our study generated a cell line with fluorescent reporters separated by a few kb only, but Gabriele *et al* finally used a theoretical closed state proxy due to the unexplainable behavior of their experimental closed state proxy. Mach *et al* used a completely colocalized cell line by targeting the same repeat array in two different colors but only used it as a verification of their closed state inference, rather than integrating it within their analysis.

Based solely on the experimental design, one could expect Mach *et al* to observe a higher proportion of contacts between anchors than Gabriele *et al* and my study because of three reasons: (1) strong synthetic CTCF triplet, (2) smaller loop size, (3) internal fluorescent reporters. This was indeed the case, as presented in **2.1.3 Comparison of results**.

2.1.2 Analysis methods

Detecting closed states

A common goal pursued by the three studies was to determine the order of magnitude of the duration and frequency of loop anchor contacts. Although this may *a priori* seem straight forward to determine from the anchor-anchor distances, we showed that this is complicated mainly by three factors: (1) stochastic motion of the chromatin polymer, (2)

finite precision in localizing fluorescent spots, (3) the non-zero genomic distance between CTCF loop anchors and the fluorescent reporters (Sabaté *et al*, NAR, 2023[101]). The three studies developed different analysis methods to segment closed states from distance time series, each with a different level of complexity. Gabriele *et al* designed a complex Bayesian inference model to infer closed states (BILD). The authors first inferred the number of switches between closed and not closed states and then inferred the closed state segmentation for each trajectory. In addition, the authors had to tune a tolerance parameter on polymer simulations to take into consideration random fluctuations in the data. Mach *et al* used a Hidden Markov Model (HMM) to segment proximal and distal states, which they trained on experimental tracks with cohesin and CTCF. Therefore, the proximal states they identify embrace the closed states with extruding states. While less complex than BILD, this analysis still relies on a model with assumptions that might not be verified in the experimental data (Gaussian distributions of the distances in closed and non-closed states). Finally, in our work, we developed a very simple closed state segmentation method by using a pair of temporal and spatial thresholds, without assuming any specific model. This method has both the advantages and disadvantages of simplicity. Due to its simplicity, our method is likely more generalizable than other more complex methods and is easily interpretable. Moreover, we did not rely on strong unverified hypotheses since our method is model-free, and likewise, we did not rely on polymer simulations to tune the detection method. However, our method cannot detect very short-lived closed states (less than ~3 minutes) and might fail for highly noisy data. These disadvantages may also occur in the two other segmentation methods.

Inferring hidden dynamic parameters of loop extrusion by comparing polymer simulations and experimental results

Gabriele *et al* and Mach *et al* used polymer simulations to explore the parameter space of cohesin processivity and density, CTCF permeability (in Gabriele *et al* only), among other parameters. By directly comparing the values obtained in simulations and in experiments, they could estimate these parameters. This approach allows to infer otherwise hidden parameters, but polymer simulations rely on models whose realism is limited. Although we try to avoid as much as possible this type of approach, they are interesting when used alongside purely experimental inferences. I briefly discuss the results obtained by this approach in the two studies below (**2.1.3 Comparison of results**), while we currently do not have estimates ourselves. Gabriele *et al* and Mach *et al* found consistent values for these

parameters as compared to previous estimations (generally inferred purely from simulations) in the literature[32,92,117,137,158].

2.1.3 Comparison of results

Cohesin constrains chromatin motion

It is not *a priori* evident whether cohesin increases or decreases chromatin motion. The dynamic process of loop extrusion could increase the motion of chromatin, as the cohesin ring 'pulls' DNA strands to form loops. On the other hand, the formation of loops might decrease chromatin motion as loops are constrained by their anchors, once extruded. All three studies revealed that cohesin-mediated loop extrusion constrains chromatin motion[163,164] (and our study). These cohesin-induced constraints on chromatin motion reduce spatial distances between DNA sequences[163].

Chromatin loops are dynamic structures

Despite the differences in the mammalian system used (mESCs *vs* human HCT116 cells), in data acquisition and in the analysis methods, the results from our three studies are mostly consistent with each other. However, several differences emerged, some of which may be explainable by simple biological arguments, while in other cases the influence of potential biological differences is difficult to disentangle from methodological differences.

All three studies found sub-hour mean durations of closed state lifetimes (15-45 min[164], 5-15 min[163], 3-23 min (our study), **Figure 20**). These estimates all argue in favor of a transient interaction between loop anchors rather than a prolonged contact. However, the fraction of time spent in the closed state was more variable between the studies. As expected from the experimental design (see above), Mach *et al* found the highest closed state fraction (27% from comparison with polymer simulations), while Gabriele *et al* found the lowest fraction (3-6.5%) and we found 8-18% and 7-27% for Tv1 and Lv4, respectively (**Figure 20**). It is noteworthy that the estimated ranges are large and future developments in the analysis of such data might help to narrow them. A similar trend was observed when assessing closed state frequencies (11.9 closed states per hour in Mach *et al*, *vs* 1 per hour in our study). Estimating absolute values from biological data, rather than relative comparisons, is difficult. Especially, Gabriele *et al* and our study had to tune the conservativeness of our closed state segmentation method without an ideal experimental ground truth. Gabriele *et al* used a theoretical proxy for

closed states and we used an experimental proxy (together with fixing the false positive rate), while Mach *et al* relied on hidden Markov models. Nevertheless, the differences between the reported results may have a biological explanation. Considering genomic distances and loop strengths (the intensity of dot interaction at TAD corners), the differences in fractions of closed states appeared to be intuitive. On the one hand, the smallest loop and strongest CTCF sites (Mach *et al*) exhibited the highest closed state fraction and frequency. On the other hand, the longer and less intense loop from Gabriele *et al* had the lowest fraction. We indeed found that Tv1 had a lower closed state fraction than Lv4, by comparing loci within our study (**Figure 20**). However, such argumentation should lead to our Tv1 loop having the lowest fraction of closed state between all studies because of its largest genomic size and lowest loop strength, which is not observed. This discrepancy has yet to be explained in terms of methodological or biological (*e.g.* human differentiated cells *vs* mouse pluripotent cells) differences.

Finally, Gabriele *et al* (by comparison with polymer simulations) and us (using experimental data) estimated the fractions of each loop state. Both our studies revealed that the active extruding state was the most represented loop state (92% in Gabriele *et al* and 76% and 43% for Lv4 and Tv1, respectively, in our study). In addition, loops were rarely observed in a purely open states (6% in Gabriele *et al*, and 0% for both Lv4 and Tv1 in our study, **Figure 20**). This analysis showed that loops were almost constantly submitted to loop extrusion, despite closed states being short-lived.

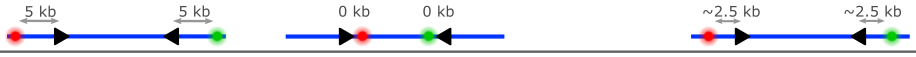
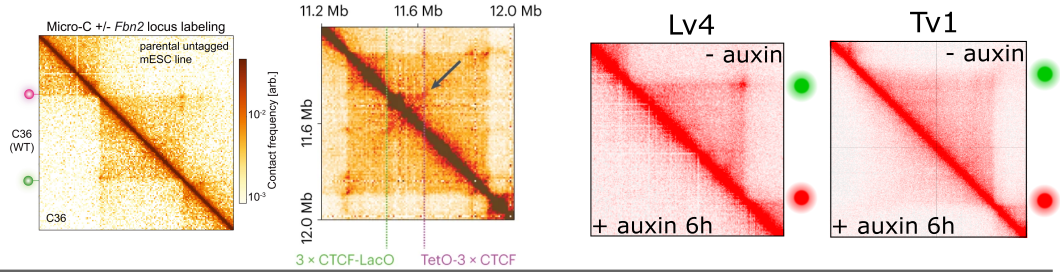
2.1.4 Discussion of results

These results highlighted the transiency of cohesin-mediated interactions between loop anchors. It is presently difficult to evaluate whether remaining quantitative differences in estimates between the studies come from biological or methodological differences. A reanalysis of the data from each study using a unique analysis method should help to disentangle their respective contribution. In addition, we used human cells instead of mESCs in the two other studies. It is possible that genome architecture dynamics, and in particular loop dynamics, is different between mouse pluripotent cells and human differentiated cells[222,280–282,309].

The mean lifetime of closed states (3-45 min from all studies together) is on the same order of magnitude as the mean residence time of cohesin molecules on chromatin (~20

min[92]). If we assume that cohesin complexes are the molecules maintaining loop anchors close to each other, it is very likely that several cohesin molecules extrude loops simultaneously. Indeed, fully extruding a loop of 300 kb would take 17 min (assuming extrusion rate of 0.3 kb/s[113]) which is already close to the mean cohesin residence time. Hence, a single cohesin molecule could not extrude fully a loop, while subsequently maintaining it in a closed state for 15 min. Similarly, the processivity of cohesin molecules was estimated at about 120-200 kb[32,164]. A single cohesin molecule is therefore unlikely to extrude a loop of 300 kb alone without detaching from chromatin. Finally, Gabriele *et al* and our study found that loops are unlikely to be found in a purely open state meaning that they are generally under active extrusion[164]. Taken together, these results argue in favor of a model where multiple cohesin molecules extrude loops simultaneously. This is also supported by the comparison of experiments and polymer simulations in Gabriele *et al* and Mach *et al* estimating extruder densities of 4.2 and 8-32 cohesin molecules per Mb, respectively (**Figure 20**). Hence, Gabriele et al estimated that ~2 cohesin molecules were simultaneously bound to the 505 kb loop they studied.

The biological implications of the dynamic and transient nature of cohesin-mediated interactions are discussed in the next sections.

	Gabriele, Brandao, Grosse-Holz et al, Science, 2022	Mach, Kos, Zhan et al, Nature Genetics, 2022	Sabaté, Preliminary results, 2023	
Genomic size	505 kb	150 kb	570 kb	925 kb
CTCF sites	Endogenous	3x strong synthetic	Endogenous	Endogenous
Labeling				
Temporal resolution	20 seconds	30 seconds	30 seconds	
Locus				
Closed Fraction	**3-6.5 %**	**27 %**	**7-27 %**	**8-18 %**
Mean Lifetime	**15-45 min**	**16 min** **5-15 min**	**3-23 min**	**4-19 min**
Frequency	**N/A**	**11.9 h^{-1}**	**1.3 h^{-1}**	**0.8 h^{-1}**
Closed	**2 %**		**24 %**	**40 %**
Extruding	**92 %**	**N/A**	**82 %**	**54 %**
Open	**6 %**		**0 %**	**0 %**
Extrusion speed	**N/A**	**N/A**	**0.23 kb/s**	**0.34 kb/s**
Extruder density	**4.2 Mb^{-1}**	**8-32 Mb^{-1}**	**N/A**	**N/A**
Cohesin processivity	**150 kb**	**170-660 kb**	**N/A**	**N/A**
Probability of CTCF to stall cohesin	**12.5 %**	**N/A**	**N/A**	**N/A**
Cohesin lifetime boost due to CTCF	**4x**	**N/A**	**N/A**	**N/A**

Experiments | **Estimates from: Comparison with polymer simulations** | **Not Measured**

Figure 20: Comparison of experimental approaches and of loop extrusion dynamics estimated from live-cell tracking of loop anchors. In the 'Labeling' row, distances between loop anchors and the closest reporter are indicated in grey. Results shown in the 'Closed fraction' and 'Closed / Extruding / Open' rows are estimated with different methods. The fraction in the 'Closed / Extruding / Open' rows may not sum to 1. In our study, ranges in closed fraction and mean lifetime indicate values estimated with more or less conservative closed state segmentation.

3.2 Cohesin molecule(s) effectively extrude loops at 0.3 kb/s

Our preliminary estimate of the effective loop extrusion speed *in vivo* is 0.3 kb/s. *In vitro* studies estimated the speed of single cohesin molecules at 0.5-1 kb/s[106,113]. An indirect estimate from dynamic Hi-C experiments evaluated the speed at ~0.38 kb/s[31]. Finally, a rough estimation based on the median size of TADs (~200 kb[30]) and the mean cohesin residence time on chromatin (~20 min[92]) yields an extrusion rate of ~0.1-0.2 kb/s. However, this last estimate ignores the fact that cohesin may remain stalled at loop anchors and is therefore likely a lower bound. The indirect and approximate extrusion rates from *in vivo* data suggest that extrusion rates are lower *in vivo* than *in vitro*. It is tempting to explain this difference by the possibility that cohesin might be decelerated by several roadbloacks during extrusion *in vivo*[107]. However, our estimate is not directly comparable to *in vitro* extrusion rates. Indeed, our estimate represents the average rate at which chromatin loops are extruded and comprises loops extruded by one or several cohesin molecules, as well as cohesins that stall before reaching the loop anchors. In addition, polymer simulations revealed that the unavoidable noise in fluorescent images tends to decrease the inferred extrusion speed (Sabaté *et al*, NAR, 2023)[101]. For these reasons, it is difficult to predict whether our estimate is an over or under estimation. Estimating the number of cohesin molecules per loop should help to compare *in vitro* and *in vivo* estimates of extrusion speeds, although the effective extrusion rate is not expected to scale linearly with the number of extruding cohesin molecules (based on our ongoing analysis of loop extrusion modeling).

Despite these arguments, our estimate is close to the previously estimated and measured extrusion speeds[31,106,113], which suggests that we were indeed able to measure the effective extrusion rate from the tracking of loop anchors. Remarkably, cohesin exhibits speeds at least one order of magnitude higher than other nuclear motors (0.03 kb/s for DNA polymerase[310], 0.025 kb/s for the RSC translocase[121] and 0.02 kb/s for RNA Pol II[120]). The high speed at which loops are extruded is likely to contribute to their dynamic nature.

3. Perspectives for future work

3.1 Towards a finer and broader quantification of chromatin loop dynamics

3.1.1 Higher accuracy in live-cell observations of chromatin loop structure

In our work, we only labelled the two loop anchors, which are the most informative loci to quantify loop dynamics, especially the closed state. However, we had to rely on analytical models to quantify the extruding state[101].

A better quantification of the extruding state, without relying on polymer or analytical models, could potentially be obtained by labeling a third locus within the loop. Using 3-color imaging, the relative distances between the two loop anchors and the third spot could be computed. Similarly to my PhD work, polymer simulations would greatly help to define the extent of information that is expected to be gained and how to quantify such data. Although spectral overlap needs to be controlled, this experiment would allow purely empirical estimate of the extruding state, as compared to the current model-based estimations[164] (and our study). Importantly, this would also require ensuring that the labeling method does not interfere with loop extrusion (*e.g*. controlling that repeat arrays do not trigger cohesin stalling).

Building up on this idea, we could label multiple (>3) loci within the loop to enable chromatin tracing in living cells. Multiplexing the colors and repeat arrays together and consequently creating a rainbow of repeat arrays within the whole loop, would allow us to trace the chromatin fiber as function of time. By mixing different types of orthogonal repeats together from blue to far-red, we could create a unique barcode for each of the loci along the chromatin fiber. Although technically challenging due to spectral overlaps, this would allow us to describe loop dynamics in much more detail.

Finally, instead of multiplexing colors, one could imagine the development of sequential imaging in living cells, similarly to sequential FISH already developed on fixed cells[134,286]. By sequentially flushing in and out detachable dyes, it would be possible to sequentially image several portions of the loop or of the genome. This is currently not possible but developments of low-affinity fluorescent probes that can be efficiently detached from their target[300] and optical developments to increase acquisition speed could allow sequential live-cell imaging with a medium temporal resolution.

Multiplexing the labelling methods could allow us to increase the accuracy at which we characterize loop structure as function of time, therefore enabling more precise and empirical estimates of loop dynamics and diminishing the need to rely on polymer models.

3.1.2 High-throughput live-cell imaging of loop anchors

The quantification of chromatin loop extrusion was performed on a few genomic loci in mouse[163,164] and human cells (our study). However, tens of thousands of loops are identified genome-wide, each harboring distinct epigenetic and transcriptional states[26,27]. It would be interesting to quantify a broad range of genomic loci to capture the variability of loop extrusion dynamics within a genome.

We tried to assess the variability of loop dynamics at dozens to hundreds of genomic loci by DNA FISH but we could not obtain sufficiently reliable signal from the images. Moreover, DNA FISH is performed on fixed cells, which cannot allow the estimation of dynamic parameters, and are subjected to fixation and denaturation artifacts. The two loci (and four in the future) that we quantified in living cells are also not sufficient to generalize our results to the entire genome. A high-throughput approach to live-cell tracking of loop anchors would be required to generalize the results from a few genomic loci.

While the generation of repair cassettes can be made at high throughput, the main difficulty comes from the generation and verification of cell lines containing fluorescently labeled loop anchors. Clone expansion and genotyping cannot be performed for one hundred different loci. Therefore, one would need to ensure that the labeling method is sufficiently specific to enable working on pools of genome edited cells without genotyping. The efficiency and specificity of labeling must be improved to ensure that a majority of cells in the pool will exhibit sufficient fluorescent signal and at the right genomic location. Indeed, depending on the targeted locus, we had as little as 10% of cell clones containing the repeat insertion at the correct locus. In a high-throughput method, a higher fraction of usable cells would be needed. Finally, high-throughput live-cell imaging methods already exist and should not cause any difficulties. Because of their theoretical ease of delivery and use, improved dCas9 systems would be the method of choice for high-throughput labeling of loop anchors in living cells.

Nevertheless, using such a high-throughput live-cell imaging method on about one hundred loops and TADs should enable to define accurately the genomic determinants of loop dynamics (euchromatin *vs* heterochromatin, gene transcription level, number and strength of CTCF site binding, *etc*.). We could also characterize the interplay between each of these genomic features (and their combination) and loop extrusion. This would allow to map the measured dynamic parameters on Hi-C maps and subsequently generalize them genome-wide by predicting loop dynamics based solely on Hi-C map patterns.

3.1.3 Cohesin pausing as function of CTCF site binding affinity

The increase in resolution of 3C techniques revealed numerous additional cohesin-dependent interactions as compared with lower resolution methods[26,27]. Specifically, multiple dot interactions were identified within loop domains. These patterns suggested that cohesin can pause at low-affinity CTCF binding sites, or other sites, during the course of extrusion. Although one of the previous studies have deleted CTCF sites within the studied loop[163], these low affinity CTCF sites exist in endogenous loops. It is therefore important to take them into consideration in our loop dynamics estimates.

It would be interesting to assess how CTCF motif binding affinity affects cohesin pausing. This can be achieved either by quantifying a high number of different genomic loci harboring various number of CTCF sites (as described above) or by studying the same locus but systematically varying the number and sequence of CTCF motifs. As above, this would require high-throughput live-cell tracking of loop anchors but is more feasible than targeting hundreds of endogenous loci since fluorescent labels only have to be inserted once.

The relationship between the number and affinity of CTCF sites, and the duration of cohesin-mediated contacts they trigger would greatly help to understand how the diversity of CTCF site distribution throughout the genome regulates chromatin loop dynamics. It would also help to better understand how evolution shaped chromatin architecture by locating and conserving clusters of CTCF sites at TAD borders[184,202].

3.1.4 Direct counting of cohesin molecules on the extruding loop in vivo

A direct readout of cohesin pausing would be to track single molecules of cohesin while simultaneously tracking CTCF sites (or the immediate flaking genomic regions) in living cells. Computing the duration of cohesin binding at CTCF sites would allow a precise

quantification of cohesin stalling. However, this is complicated by two main reasons: (1) the optical resolution which limits the ability to decipher whether cohesin is bound on the visualized CTCF site or on a nearby genomic region, (2) the low labeling density required to perform single-particle tracking (SPT). In usual SPT, single molecules can be visualized only if a minor fraction of the total molecules are fluorescently labelled. Hence, a non-fluorescent molecule of cohesin is more likely to bind on the visualized CTCF site than a fluorescent one, drastically increasing the imaging duration to record a sufficient number of events.

We designed our cell lines in a way that could allow such an imaging approach. We labelled RAD21 with a SNAP tag that could allow SPT of cohesin molecules simultaneously with loop anchor imaging in living cells. This could allow us to visualize the cohesin molecule(s) actively extruding the visualized loop. However, based on polymer and RAD21 molecule diffusion simulations, we found that, although not impossible, this plan is extremely ambitious technically for the reasons explained above. By simulating microscopy images from polymer simulations, we defined that labeling a maximum of 10-20% of RAD21 molecules should allow us to distinguish bound molecules with lattice light-sheet microscopy. But given the rarity of closed states, the lack of resolution and the sparsity of labeling, closed state events with a fluorescent cohesin molecule may still be too rare to be imaged. However, using localized photoactivation of cohesin molecules around the visualized loop and real-time monitoring of loop anchor distance, and by triggering RAD21 SPT imaging when detecting a decrease in loop anchor distance, one could visualize such events with sufficient likelihood. Although of great interest, this remains an extremely challenging approach that would require considerable further work.

3.1.5 Reorganization of chromatin architecture after mitosis

In our study, we focused on the dynamics of loop anchors in the G1 phase of the cell cycle. However, genome architecture undergoes a major change during cell division. From Hi-C data of synchronized cell populations, it was shown that chromatin loops already appear as soon as in anaphase/telophase and are further reinforced until mid-G1[311–313]. Using our loop anchor labelled cell lines and by synchronizing cells, it would be possible to quantify at unprecedent temporal resolution the re-formation of chromatin loops after mitotic exit and the duration needed to reach G1 steady-state in loop anchor distance. By depleting RAD21 from the cells, it would be possible to identify to what extent cohesin molecules contribute to

chromatin architecture re-formation after mitotic exit, as compared to other mechanisms such as A/B compartmentalization.

3.2 Functional implications of the dynamic nature of chromatin loops

Our work on quantitatively estimating dynamic parameters of chromatin loops remains insignificant if not put in biological context. Indeed, from our and others results, we conclude that the biological function of loop extrusion is to transiently increase the frequency of long-range interactions, at many sites within the loop, rather than creating prolonged contacts between loop anchors. However, the key functional consequences appear when considering how these transient interactions affect the major nuclear processes such as gene expression regulation or DNA repair. Therefore, based on our cell lines, several experimental ideas could be conducted to analyze in more detail how loop extrusion contributes to proper storage, maintenance and expression of the genome.

3.2.1 Transient Enhancer-Promoter contacts as a gene expression model?

Although contradictory results exist (see **4.2 Loop extrusion can help to regulate gene expression**), it has been proposed that cohesin-mediated contacts could allow proper gene regulation by bringing enhancer and promoters together, or on the contrary by impeding enhancer-promoter contacts across TAD boundaries[138,210]. Our findings that loop extrusion creates transient rather than prolonged contacts are consistent with the observed stochasticity of gene expression[314,315] and could potentially help build models of gene expression activation. Gene expression activates through the binding of general and specific transcription factors (TF). However, the exact sequence and duration of transcription factor binding before gene activation is still unclear. By mediating long-range interactions, loop extrusion might allow TFs to contact promoters that are spatially away from their enhancer(s). The transient contacts created by loop extrusion and the importance of low-affinity interactions in TF binding[316] argue in favor of a 'hit and run' model of gene activation[317,318]. Multiple and frequent, but short-lived TF binding could help explain stochasticity in gene expression[315,319]. The assembly of molecular factors necessary for gene activation might only use transient and frequent interactions rather than a sequentially ordered sequence of prolonged binding events. However, based on the frequency of closed states (1.2 per hour), it is likely that loop extrusion is not the only mechanism allowing enhancer-promoter contacts and that additional specific interactions are involved for gene regulation.

In this work, I chose to exclude chromatin loops containing expressed genes from our set of labelled loops to avoid obscuring extrusion-dependent motion of chromatin. However, it would be interesting to label loop anchors, as well as to follow the transcriptional output by MS2 labelling of the RNA[320] using a third color. The presence or absence of correlation between loop anchor distance and RNA production could be assessed to decipher the role of loop extrusion in gene expression[244,321] Due to the high diversity of gene regulation, only the study of a significant number of endogenous loci, or a carefully designed study, would provide a consistent answer.

3.2.2 Loop extrusion contributes to proper DNA repair

3D chromatin organization and loop extrusion have been proposed to play a role in the correct repair of DNA breaks[246,247,253]. However, polymer simulations suggested that plausible scenarios of DNA repair involving loop extrusion require large changes in loop extrusion dynamics (**Figure 13C**). In these scenarios, cohesin processivity would be largely increased and cohesin would extrude loops for long periods of time (**Figure 13C**)[253]. Therefore, the characterized dynamic nature of loop extrusion in G1 cell[163,164] (and our study) might be tuned and modified to achieve different biological functions, eventually using cohesin variants such as SMC5/6 instead of SMC1/3[322].

To clarify the mechanism by which loop extrusion contributes to DNA repair[247] and test theoretical predictions[253], our cell lines could be used in two different ways. First, our parental cell line with the auxin-dependent RAD21-mAID-SNAP fusion would allow to: (1) quantify RAD21 binding with or without DNA damage to test theoretical predictions of increased lifetime[253] by SPT, (2) assess whether RAD21 is preferentially recruited or passively accumulates at DSB sites by local laser-induced DNA damage followed by RAD21 SPT, and (3) study the spatial organization of γH2AX foci relative to cohesin by 2-color super-resolution imaging: cohesin might be found at the edge of γH2AX foci if it is involved in the deposition of the histone marks. Second, by adding rare restriction enzyme sites to induce DNA breaks[323] at the labelled loop anchors or at other sites in the genome, it would become possible to visualize and quantify how chromatin loop dynamics are affected upon DNA damage. Because cohesin might play a role in sister chromatid cohesion during DNA repair, the results of such studies need to be carefully interpreted to decipher whether the loop extrusion activity or the cohesion activity of cohesin are measured.

4. Loop extrusion allows a large range of interactions to occur without determining their outcome

We and others showed that loop extrusion creates short-lived long-range chromatin interactions[163,164]. The loop extrusion process *per se* is extremely simple: cohesin randomly binds on DNA, extrudes loops and unbinds. However, a complex and precise regulation of this process arose throughout evolution, with multiple molecules (DNA, RNA and proteins) achieving the tight regulation of this process. Controlling the dynamics of loop extrusion and thereby functionally important chromatin interactions is likely of crucial importance for the cell.

DNA sequences that are located away from each other on the genome sequence have a low probability of contacting each other due to passive motion of the chromatin only. Loop extrusion enables such contacts, transiently. Therefore, contacts that have a biological function, such as EP contacts, are made possible thanks to loop extrusion and could be reinforced by specific interactions upon encounter. However, unspecific contacts that would be detrimental for the cell such as between an enhancer and an oncogene will be limited by the transiency of cohesin-mediated interactions. If extrusion had generated prolonged contacts, the probability of inducing such detrimental interactions would have increased. Therefore, the transient and dynamic nature of cohesin-mediated interactions may be a means to prevent undesired functional consequences of chromatin interactions.

Thus, by creating short-lived long-range interactions, loop extrusion dynamics allows specific contacts to occur, while preventing robust and prolonged, potentially detrimental DNA interactions. We can speculate that robustness of interactions is provided by the sum of low-affinity interactions that allows specific associations to happen despite the dynamic and transient nature of cohesin-mediated contacts.

References

1. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
2. Mack, A. H., Schlingman, D. J., Ilagan, R. P., Regan, L. & Mochrie, S. G. J. Kinetics and Thermodynamics of Phenotype: Unwinding and Rewinding the Nucleosome. *Journal of Molecular Biology* **423**, 687–701 (2012).
3. Yan, J. *et al.* Micromanipulation Studies of Chromatin Fibers in Xenopus Egg Extracts Reveal ATP-dependent Chromatin Assembly Dynamics. *MBoC* **18**, 464–474 (2007).
4. Lai, W. K. M. & Pugh, B. F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol* **18**, 548–562 (2017).
5. Kireeva, M. L. *et al.* Nature of the Nucleosomal Barrier to RNA Polymerase II. *Molecular Cell* **18**, 97–108 (2005).
6. Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
7. Zhang, Z. *et al.* A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome. *Science* **332**, 977–980 (2011).
8. Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell* **154**, 490–503 (2013).
9. Rando, O. J. & Chang, H. Y. Genome-wide views of chromatin structure. *Annu Rev Biochem* **78**, 245–271 (2009).
10. Turner, B. M. Defining an epigenetic code. *Nat Cell Biol* **9**, 2–6 (2007).
11. Turner, B. M. Histone acetylation and an epigenetic code. *BioEssays* **22**, 836–845 (2000).
12. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications – writers that read. *EMBO Rep* **16**, 1467–1481 (2015).

13. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**, 2478–2492 (2017).

14. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381–395 (2011).

15. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences* **35**, 618–626 (2010).

16. Almouzni, G. & Cedar, H. Maintenance of Epigenetic Information. *Cold Spring Harb Perspect Biol* **8**, a019372 (2016).

17. Heard, E. & Martienssen, R. A. Transgenerational Epigenetic Inheritance: myths and mechanisms. *Cell* **157**, 95–109 (2014).

18. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487–500 (2016).

19. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2018).

20. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Science Advances* **5**, eaaw1668 (2019).

21. Akgol Oksuz, B. *et al.* Systematic evaluation of chromosome conformation capture assays. *Nat Methods* **18**, 1046–1055 (2021).

22. Belton, J.-M. *et al.* Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).

23. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat Rev Genet* **21**, 207–226 (2019).

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

25. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).

26. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell* **78**, 554-565.e7 (2020).

27. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell* **78**, 539-553.e8 (2020).

28. Goel, V. Y. & Hansen, A. S. The macro and micro of chromosome conformation capture. *WIREs Developmental Biology* **10**, e395 (2021).

29. Hua, P. *et al.* Defining genome architecture at base-pair resolution. *Nature* **595**, 125–129 (2021).

30. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

31. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).

32. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* **15**, 2038–2049 (2016).

33. Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile pile-up analysis of Hi-C data. *Bioinformatics* **36**, 2980–2985 (2020).

34. Scolari, V. F., Mercy, G., Koszul, R., Lesne, A. & Mozziconacci, J. Kinetic Signature of Cooperativity in the Irreversible Collapse of a Polymer. *Phys. Rev. Lett.* **121**, 057801 (2018).

35. Fudenberg, G. & Imakaev, M. FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat Methods* **14**, 673–678 (2017).

36. Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biology* **17**, 215 (2016).

37. Traganos, F., Darzyndiewicz, Z., Sharpless, T. & Melamed, M. R. Denaturation of deoxyribonucleic acid in situ effect of formaldehyde. *J Histochem Cytochem* **23**, 431–438 (1975).

38. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

39. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).

40. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757.e24 (2018).

41. Fiorillo, L. *et al.* Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin. *Nat Methods* **18**, 482–490 (2021).

42. Rosa, A. & Zimmer, C. Chapter Nine - Computational Models of Large-Scale Genome Architecture. in *International Review of Cell and Molecular Biology* (eds. Hancock, R. & Jeon, K. W.) vol. 307 275–349 (Academic Press, 2014).

43. Sazer, S. & Schiessel, H. The biology and polymer physics underlying large□scale chromosome organization. *Traffic* **19**, 87–104 (2018).

44. Mirny, L. A., Imakaev, M. & Abdennur, N. Two major mechanisms of chromosome organization. *Current Opinion in Cell Biology* **58**, 142–152 (2019).

45. Podgornik, R. Statistical physics of macromolecules. *J Stat Phys* **78**, 1179–1180 (1995).

46. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res* **19**, 37–51 (2011).

47. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol* **30**, 90–98 (2012).

48. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).

49. Barbieri, M. *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences* **109**, 16173–16178 (2012).

50. Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer physics of chromosome large-scale 3D organisation. *Sci Rep* **6**, 29775 (2016).

51. Briane, V., Kervrann, C. & Vimond, M. Statistical analysis of particle trajectories in living cells. *Phys Rev E* **97**, 062121 (2018).

52. Backlund, M. P., Joyner, R. & Moerner, W. E. Chromosomal locus tracking with proper accounting of static and dynamic errors. *Phys Rev E Stat Nonlin Soft Matter Phys* **91**, 062716 (2015).

53. Hajjoul, H. *et al.* High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res.* **23**, 1829–1838 (2013).

54. Keizer, V. I. P. *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science* **377**, 489–495 (2022).

55. Chiang, M., Brackley, C. A., Marenduzzo, D. & Gilbert, N. Predicting genome organisation and function with mechanistic modelling. *Trends in Genetics* **0**, (2021).

56. Esposito, A. *et al.* Polymer models are a versatile tool to study chromatin 3D organization. *Biochem Soc Trans* **49**, 1675–1684 (2021).

57. Imakaev, M. V., Fudenberg, G. & Mirny, L. A. Modeling chromosomes: Beyond pretty pictures. *FEBS Lett.* **589**, 3031–3036 (2015).

58. Conte, M. *et al.* Loop-extrusion and polymer phase-separation can co-exist at the single-molecule level to shape chromatin folding. *Nat Commun* **13**, 4070 (2022).

59. Conte, M. *et al.* Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nat Commun* **11**, 1–13 (2020).

60. Rosa, A. & Everaers, R. Structure and Dynamics of Interphase Chromosomes. *PLOS Computational Biology* **4**, e1000153 (2008).

61. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6697–E6706 (2018).

62. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6456-6465 (2015).

63. Giorgetti, L. *et al.* Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription. *Cell* **157**, 950–963 (2014).

64. Baù, D. *et al.* The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol* **18**, 107–114 (2011).

65. Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3D genome reconstruction from chromosomal contacts. *Nat Methods* **11**, 1141–1143 (2014).

66. Hu, M. *et al.* Bayesian Inference of Spatial Organizations of Chromosomes. *PLOS Computational Biology* **9**, e1002893 (2013).

67. Zhang, Z., Li, G., Toh, K.-C. & Sung, W.-K. 3D Chromosome Modeling with Semi-Definite Programming and Hi-C Data. *Journal of Computational Biology* **20**, 831–846 (2013).

68. Rajpurkar, A. R., Mateo, L. J., Murphy, S. E. & Boettiger, A. N. Deep learning connects DNA traces to transcription to reveal predictive features beyond enhancer–promoter contact. *Nat Commun* **12**, 3423 (2021).

69. Parmar, J. J., Woringer, M. & Zimmer, C. How the Genome Folds: The Biophysics of Four-Dimensional Chromatin Organization. *Annu. Rev. Biophys.* **48**, 231–253 (2019).

70. Naumova, N. *et al.* Organization of the Mitotic Chromosome. *Science* **342**, 948–953 (2013).

71. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb Perspect Biol* **2**, a003889 (2010).

72. Bolzer, A. *et al.* Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. *PLOS Biology* **3**, e157 (2005).

73. Girelli, G. *et al.* GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat Biotechnol* **38**, 1184–1193 (2020).

74. Alberts, B. *et al. Molecular Biology of the Cell*. (Garland Science, 2002).

75. Mirny, L. & Dekker, J. Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. *Cold Spring Harb Perspect Biol* **14**, a040147 (2021).

76. Spracklin, G. *et al.* Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. *Nat Struct Mol Biol* **30**, 38–51 (2023).

77. Nichols, M. H. & Corces, V. G. Principles of 3D compartmentalization of the human genome. *Cell Reports* **35**, 109330 (2021).

78. Fraser, J. *et al.* Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular Systems Biology* **11**, 852 (2015).

79. Liu, Y. *et al.* Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nat Commun* **12**, 2439 (2021).

80. Yildirim, A., Boninsegna, L., Zhan, Y. & Alber, F. Uncovering the Principles of Genome Folding by 3D Chromatin Modeling. *Cold Spring Harb Perspect Biol* **14**, a039693 (2022).

81. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).

82. Erdel, F. & Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal* **114**, 2262–2270 (2018).

83. Liu, X. *et al.* Time-dependent effect of 1,6-hexanediol on biomolecular condensates and 3D chromatin organization. *Genome Biol* **22**, 230 (2021).

84. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).

85. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

86. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

87. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).

88. Hsieh, T.-H. S. *et al.* Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat Genet* **54**, 1919–1932 (2022).

89. Szabo, Q. *et al.* TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Science Advances* **4**, eaar8082 (2018).

90. Szabo, Q. *et al.* Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics* **52**, 1151–1157 (2020).

91. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **173**, 1165-1178.e20 (2018).

92. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6**, e25776 (2017).

93. Gu, B. *et al.* Opposing Effects of Cohesin and Transcription on CTCF Organization Revealed by Super-resolution Imaging. *Molecular Cell* **80**, 699-711.e7 (2020).

94. Li, Y. *et al.* The structural basis for cohesin–CTCF-anchored loops. *Nature* **578**, 472–476 (2020).

95. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *The EMBO Journal* **36**, 3573–3599 (2017).

96. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).

97. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).

98. Davidson, I. F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* **22**, 445–464 (2021).

99. Hansen, A. S. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* **11**, 132–148 (2020).

100. Anania, C. *et al.* In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. *Nat Genet* **54**, 1026–1036 (2022).

101. Sabaté, T., Lelandais, B., Bertrand, E. & Zimmer, C. Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging. *Nucleic Acids Research* **51**, 2614–2632 (2023).

102. Guo, Y. *et al.* Chromatin jets define the properties of cohesin-driven in vivo loop extrusion. *Molecular Cell* **82**, 3769-3780.e5 (2022).

103. Wike, C. L. *et al.* Chromatin architecture transitions from zebrafish sperm through early embryogenesis. *Genome Res.* **31**, 981–994 (2021).

104. Kim, E., Barth, R. & Dekker, C. Looping the Genome with SMC Complexes. *Annual Review of Biochemistry* **92**, 15–41 (2023).

105. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.* **82**, 45–55 (2017).

106. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).

107. Pradhan, B. *et al.* SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Reports* **41**, (2022).

108. Barth, R. *et al.* Testing pseudotopological and nontopological models for SMC-driven DNA loop extrusion against roadblock-traversal experiments. *Sci Rep* **13**, 8100 (2023).

109. Srinivasan, M. *et al.* The Cohesin Ring Uses Its Hinge to Organize DNA Using Non-topological as well as Topological Mechanisms. *Cell* **173**, 1508-1519.e18 (2018).

110. Banigan, E. J., van den Berg, A. A., Brandão, H. B., Marko, J. F. & Mirny, L. A. Chromosome organization by one-sided and two-sided loop extrusion. *eLife* **9**, e53558 (2020).

111. Banigan, E. J. & Mirny, L. A. Limits of Chromosome Compaction by Loop-Extruding Motors. *Phys. Rev. X* **9**, 031007 (2019).

112. Banigan, E. J. & Mirny, L. A. The interplay between asymmetric and symmetric DNA loop extrusion. *eLife* **9**, e63528 (2020).

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

114. Liu, N. Q. *et al. Rapid depletion of CTCF and cohesin proteins reveals dynamic features of chromosome architecture*. 2021.08.27.457977 https://www.biorxiv.org/content/10.1101/2021.08.27.457977v1 (2021) doi:10.1101/2021.08.27.457977.

115. Matityahu, A. & Onn, I. It's all in the numbers: Cohesin stoichiometry. *Frontiers in Molecular Biosciences* **9**, (2022).

116. Liu, W. *et al.* Monomeric cohesin state revealed by live-cell single-molecule spectroscopy. *EMBO reports* **21**, e48211 (2020).

117. Cattoglio, C. *et al.* Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *eLife* **8**, e40164 (2019).

118. Nomidis, S. K., Carlon, E., Gruber, S. & Marko, J. F. DNA tension-modulated translocation and loop extrusion by SMC complexes revealed by molecular dynamics simulations. *Nucleic Acids Res* **50**, 4974–4987 (2022).

119. Banigan, E. J. & Mirny, L. A. Loop extrusion: theory meets single-molecule experiments. *Current Opinion in Cell Biology* **64**, 124–138 (2020).

120. Muniz, L., Nicolas, E. & Trouche, D. RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *EMBO J* **40**, e105740 (2021).

121. Sirinakis, G. *et al.* The RSC chromatin remodelling ATPase translocates DNA with high force and small step size. *EMBO J* **30**, 2364–2372 (2011).

122. Rippe, K. Liquid–Liquid Phase Separation in Chromatin. *Cold Spring Harb Perspect Biol* **14**, a040683 (2022).

123. Kim, E., Kerssemakers, J., Shaltiel, I. A., Haering, C. H. & Dekker, C. DNA-loop extruding condensin complexes can traverse one another. *Nature* **579**, 438–442 (2020).

124. Brandão, H. B., Ren, Z., Karaboja, X., Mirny, L. A. & Wang, X. DNA-loop-extruding SMC complexes can traverse one another in vivo. *Nat Struct Mol Biol* **28**, 642–651 (2021).

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

126. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell* **60**, 676–684 (2015).

127. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–910 (2015).

128. Kraft, K. *et al.* Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat Cell Biol* **21**, 305–310 (2019).

129. Segueni, J. & Noordermeer, D. CTCF: A misguided jack-of-all-trades in cancer cells. *Comput Struct Biotechnol J* **20**, 2685–2698 (2022).

130. Nora, E. P. *et al.* Molecular basis of CTCF binding polarity in genome folding. *Nature Communications* **11**, 5612 (2020).

131. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).

132. Finn, E. H. *et al.* Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502-1515.e10 (2019).

133. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).

134. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).

135. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49 (2019).

136. Chang, L.-H., Ghosh, S. & Noordermeer, D. TADs and their borders: free movement or building a wall? *Journal of Molecular Biology* **432**, 643–652 (2020).

137. Davidson, I. F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822–827 (2023).

138. Xiao, J. Y., Hafner, A. & Boettiger, A. N. How subtle changes in 3D structure can create large changes in transcription. *eLife* **10**, e64320 (2021).

139. Luppino, J. M. *et al.* Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nature Genetics* **52**, 840–848 (2020).

140. Hafner, A. *et al.* Loop stacking organizes genome folding from TADs to chromosomes. *Molecular Cell* **83**, 1377-1392.e6 (2023).

141. Anderson, E. C. & Nora, E. P. Setting new boundaries with transcription and CTCF. *Nat Genet* **52**, 1003–1004 (2020).

142. Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503–507 (2017).

143. Banigan, E. J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proceedings of the National Academy of Sciences* **120**, e2210480120 (2023).

144. Heinz, S. *et al.* Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**, 1522-1536.e22 (2018).

145. Brandão, H. B. *et al.* RNA polymerases as moving barriers to condensin loop extrusion. *PNAS* **116**, 20489–20499 (2019).

146. Neguembor, M. V. *et al.* Transcription-mediated supercoiling regulates genome folding and loop formation. *Molecular Cell* **81**, 3065-3091.e12 (2021).

147. Hansen, A. S. *et al.* Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Molecular Cell* **76**, 395-411.e13 (2019).

148. Dequeker, B. J. H. *et al.* MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature* **606**, 197–203 (2022).

149. Jeppsson, K. *et al.* Cohesin-dependent chromosome loop extrusion is limited by transcription and stalled replication forks. *Sci Adv* **8**, eabn7063 (2022).

150. Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).

151. Uusküla-Reimand, L. *et al.* Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* **17**, 182 (2016).

152. Canela, A. *et al.* Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. *Molecular Cell* **75**, 252-266.e8 (2019).

153. Canela, A. *et al.* Genome Organization Drives Chromosome Fragility. *Cell* **170**, 507-521.e18 (2017).

154. Kagey, M. H. *et al.* Mediator and Cohesin Connect Gene Expression and Chromatin Architecture. *Nature* **467**, 430–435 (2010).

155. Haarhuis, J. H. I. *et al.* A Mediator-cohesin axis controls heterochromatin domain formation. *Nat Commun* **13**, 754 (2022).

156. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).

157. Conte, M., Esposito, A., Vercellone, F., Abraham, A. & Bianco, S. Unveiling the Machinery behind Chromosome Folding by Polymer Physics Modeling. *International Journal of Molecular Sciences* **24**, 3660 (2023).

158. Holzmann, J. *et al.* Absolute quantification of cohesin, CTCF and their regulators in human cells. *eLife* **8**, e46269 (2019).

159. Cardozo Gizzi, A. M. *et al.* Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell* **74**, 212-222.e5 (2019).

160. Mateo, L. J., Sinnott-Armstrong, N. & Boettiger, A. N. Tracing DNA paths and RNA profiles in cultured cells and tissues with ORCA. *Nature Protocols* **16**, 1647–1713 (2021).

161. Sood, V. & Misteli, T. The stochastic nature of genome organization and function. *Current Opinion in Genetics & Development* **72**, 45–52 (2022).

162. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).

163. Mach, P. *et al.* Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet* **54**, 1907–1918 (2022).

164. Gabriele, M. *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496–501 (2022).

165. Cuadrado, A. & Losada, A. Specialized functions of cohesins STAG1 and STAG2 in 3D genome architecture. *Current Opinion in Genetics & Development* **61**, 9–16 (2020).

166. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707.e14 (2017).

167. Liu, N. Q. *et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nature Genetics* **53**, 100–109 (2021).

168. Morales, C. *et al.* PDS5 proteins are required for proper cohesin dynamics and participate in replication fork protection. *J. Biol. Chem.* **295**, 146–157 (2020).

169. Kanke, M., Tahara, E., Huis in't Veld, P. J. & Nishiyama, T. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *The EMBO Journal* **35**, 2686–2698 (2016).

170. Luppino, J. M. *et al.* Co-depletion of NIPBL and WAPL balance cohesin activity to correct gene misexpression. *PLOS Genetics* **18**, e1010528 (2022).

171. Shin, H. & Kim, Y. Regulation of loop extrusion on the interphase genome. *Critical Reviews in Biochemistry and Molecular Biology* **0**, 1–18 (2023).

172. Alomer, R. M. *et al.* Esco1 and Esco2 regulate distinct cohesin functions during cell cycle progression. *Proceedings of the National Academy of Sciences* **114**, 9906–9911 (2017).

173. Kawasumi, R. *et al.* ESCO1/2's roles in chromosome structure and interphase chromatin organization. *Genes Dev.* **31**, 2136–2150 (2017).

174. Minamino, M. *et al.* Esco1 Acetylates Cohesin via a Mechanism Different from That of Esco2. *Current Biology* **25**, 1694–1706 (2015).

175. Ben-Shahar, T. R. *et al.* Eco1-Dependent Cohesin Acetylation During Establishment of Sister Chromatid Cohesion. *Science* **321**, 563–566 (2008).

176. Ünal, E. *et al.* A Molecular Determinant for the Establishment of Sister Chromatid Cohesion. *Science* **321**, 566–569 (2008).

177. Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M. & Ellenberg, J. Live-Cell Imaging Reveals a Stable Cohesin-Chromatin Interaction after but Not before DNA Replication. *Current Biology* **16**, 1571–1578 (2006).

178. Wutz, G. *et al.* ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. *Elife* **9**, (2020).

179. Schmitz, J., Watrin, E., Lénárt, P., Mechtler, K. & Peters, J.-M. Sororin Is Required for Stable Binding of Cohesin to Chromatin and for Sister Chromatid Cohesion in Interphase. *Current Biology* **17**, 630–636 (2007).

180. Losada, A. Cohesin in cancer: chromosome segregation and beyond. *Nat Rev Cancer* **14**, 389–393 (2014).

181. Morales, C. & Losada, A. Establishing and dissolving cohesion during the vertebrate cell cycle. *Curr Opin Cell Biol* **52**, 51–57 (2018).

182. van Ruiten, M. S. *et al.* The cohesin acetylation cycle controls chromatin loop length through a PDS5A brake mechanism. *Nat Struct Mol Biol* **29**, 586–591 (2022).

183. Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **50**, D165–D173 (2022).

184. Kentepozidou, E. *et al.* Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biology* **21**, 5 (2020).

185. Huang, H. *et al.* CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nature Genetics* **53**, 1064–1074 (2021).

186. Willemin, A. *et al.* Induction of a chromatin boundary in vivo upon insertion of a TAD border. *PLOS Genetics* **17**, e1009691 (2021).

187. Luan, J. *et al.* Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. *Cell Reports* **34**, 108783 (2021).

188. Marina-Zárate, E., Rodríguez-Ronchel, A., Gómez, M. J., Sánchez-Cabo, F. & Ramiro, A. R. Low-affinity CTCF binding drives transcriptional regulation whereas high-affinity binding encompasses architectural functions. *iScience* **26**, 106106 (2023).

189. Holwerda, S. J. B. & de Laat, W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120369 (2013).

190. Hashimoto, H. *et al.* Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Molecular Cell* **66**, 711-720.e3 (2017).

191. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* **22**, 1680–1688 (2012).

192. Fudenberg, G., Kelley, D. R. & Pollard, K. S. Predicting 3D genome folding from DNA sequence with Akita. *Nature Methods* **17**, 1111–1117 (2020).

193. Schwessinger, R. *et al.* DeepC: predicting 3D genome folding using megabase-scale transfer learning. *Nat Methods* **17**, 1118–1124 (2020).

194. Chang, L.-H. *et al.* A complex CTCF binding code defines TAD boundary structure and function. *bioRxiv* 2021.04.15.440007 (2021) doi:10.1101/2021.04.15.440007.

195. Nanni, L., Ceri, S. & Logie, C. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biology* **21**, 197 (2020).

196. Hansen, A. S., Amitai, A., Cattoglio, C., Tjian, R. & Darzacq, X. Guided nuclear exploration increases CTCF target search efficiency. *Nat Chem Biol* **16**, 257–266 (2020).

197. Kung, J. T. *et al.* Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. *Mol. Cell* **57**, 361–375 (2015).

198. Saldaña-Meyer, R. *et al.* RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Molecular Cell* **76**, 412-422.e5 (2019).

199. Luo, H. *et al.* HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia. *Molecular Cell* **82**, 833-851.e11 (2022).

200. Zhang, H. *et al.* CTCF and R-loops are boundaries of cohesin-mediated DNA looping. 2022.09.15.508177 Preprint at https://doi.org/10.1101/2022.09.15.508177 (2022).

201. Tsai, P.-F. *et al.* A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. *Molecular Cell* **71**, 129-141.e8 (2018).

202. Álvarez-González, L. *et al.* Principles of 3D chromosome folding and evolutionary genome reshuffling in mammals. *Cell Reports* **41**, 111839 (2022).

203. Li, D. *et al.* Comparative 3D genome architecture in vertebrates. *BMC Biology* **20**, 99 (2022).

204. Corbo, M., Damas, J., Bursell, M. G. & Lewin, H. A. Conservation of chromatin conformation in carnivores. *PNAS* **119**, (2022).

205. Choudhary, M. N. K., Quaid, K., Xing, X., Schmidt, H. & Wang, T. Widespread contribution of transposable elements to the rewiring of mammalian 3D genomes. *Nat Commun* **14**, 634 (2023).

206. Cheng, H., Zhang, N. & Pati, D. Cohesin subunit RAD21: From biology to disease. *Gene* **758**, 144966 (2020).

207. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e18 (2018).

208. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).

209. Oti, M., Falck, J., Huynen, M. A. & Zhou, H. CTCF-mediated chromatin loops enclose inducible gene regulatory domains. *BMC Genomics* **17**, 252 (2016).

210. Karpinska, M. A. & Oudelaar, A. M. The role of loop extrusion in enhancer-mediated gene activation. *Current Opinion in Genetics & Development* **79**, 102022 (2023).

211. Rinzema, N. J. *et al.* Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nat Struct Mol Biol* **29**, 563–574 (2022).

212. Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Cell* **184**, 5775-5790.e30 (2021).

213. Calderon, L. *et al.* Cohesin-dependence of neuronal gene expression relates to chromatin loop length. *eLife* **11**, e76539 (2022).

214. Cuartero, S. *et al.* Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat Immunol* **19**, 932–941 (2018).

215. Robles-Rebollo, I. *et al.* Cohesin couples transcriptional bursting probabilities of inducible enhancers and promoters. *Nat Commun* **13**, 4342 (2022).

216. Stik, G. *et al.* CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. *Nat Genet* **52**, 655–661 (2020).

217. Grubert, F. *et al.* Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737–743 (2020).

218. Plaza-Jennings, A. L. *et al.* HIV integration in the human brain is linked to microglial activation and 3D genome remodeling. *Molecular Cell* **82**, 4647-4663.e8 (2022).

219. Wang, C. *et al.* A DNA tumor virus globally reprograms host 3D genome architecture to achieve immortal growth. *Nat Commun* **14**, 1598 (2023).

220. Wang, R. *et al.* SARS-CoV-2 restructures host chromatin architecture. *Nat Microbiol* **8**, 679–694 (2023).

221. Greenwald, W. W. *et al.* Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. *Nat Commun* **10**, 1054 (2019).

222. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).

223. Winick-Ng, W. *et al.* Cell-type specialization is encoded by specific chromatin topologies. *Nature* **599**, 684–691 (2021).

224. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).

225. Kane, L. *et al.* Cohesin is required for long-range enhancer action at the Shh locus. *Nat Struct Mol Biol* **29**, 891–897 (2022).

226. Thiecke, M. J. *et al.* Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Reports* **32**, 107929 (2020).

227. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).

228. Jia, Z. *et al.* Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biology* **21**, 75 (2020).

229. Aljahani, A. *et al.* Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat Commun* **13**, 2139 (2022).

230. Zuin, J. *et al.* Nonlinear control of transcription through enhancer–promoter interactions. *Nature* **604**, 571–577 (2022).

231. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).

232. Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016).

233. Cavalheiro, G. R., Pollex, T. & Furlong, E. E. To loop or not to loop: what is the role of TADs in enhancer function and gene regulation? *Current Opinion in Genetics & Development* **67**, 119–129 (2021).

234. Amândio, A. R. *et al.* Sequential in cis mutagenesis in vivo reveals various functions for CTCF sites at the mouse HoxD cluster. *Genes Dev.* **35**, 1490–1509 (2021).

235. Rodríguez-Carballo, E. *et al.* Chromatin topology and the timing of enhancer function at the HoxD locus. *Proc Natl Acad Sci U S A* **117**, 31231–31241 (2020).

236. Rodríguez-Carballo, E., Lopez-Delisle, L., Yakushiji-Kaminatsui, N., Ullate-Agote, A. & Duboule, D. Impact of genome architecture on the functional activation and repression of Hox regulatory landscapes. *BMC Biology* **17**, 55 (2019).

237. Rodríguez-Carballo, E. *et al.* The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).

238. Paliou, C. *et al.* Preformed chromatin topology assists transcriptional robustness of *Shh* during limb development. *Proceedings of the National Academy of Sciences* **116**, 12390–12399 (2019).

239. Despang, A. *et al.* Functional dissection of the Sox9 – Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet* **51**, 1263–1271 (2019).

240. Williamson, I. *et al.* Developmentally regulated Shh expression is robust to TAD perturbations. *Development* **146**, dev179523 (2019).

241. Chakraborty, S. *et al.* Enhancer–promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. *Nat Genet* **55**, 280–290 (2023).

242. Williamson, I., Lettice, L. A., Hill, R. E. & Bickmore, W. A. Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* **143**, 2994–3001 (2016).

243. Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Molecular Cell* **76**, 473-484.e7 (2019).

244. Alexander, J. M. *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* **8**, e41769 (2019).

245. Rekaik, H. *et al.* Sequential and directional insulation by conserved CTCF sites underlies the Hox timer in stembryos. *Nat Genet* 1–12 (2023) doi:10.1038/s41588-023-01426-7.

246. Kalousi, A. & Soutoglou, E. Nuclear compartmentalization of DNA repair. *Current Opinion in Genetics & Development* **37**, 148–157 (2016).

247. Arnould, C. *et al.* Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660–665 (2021).

248. Fu, J. *et al.* ATM–ESCO2–SMC3 axis promotes 53BP1 recruitment in response to DNA damage and safeguards genome integrity by stabilizing cohesin complex. *Nucleic Acids Research* gkad533 (2023) doi:10.1093/nar/gkad533.

249. Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase Chromatin Domains Involved in DNA Double-Strand Breaks in Vivo. *Journal of Cell Biology* **146**, 905–916 (1999).

250. Mah, L.-J., El-Osta, A. & Karagiannis, T. C. γH2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia* **24**, 679–686 (2010).

251. Phipps, J. & Dubrana, K. DNA Repair in Space and Time: Safeguarding the Genome with the Cohesin Complex. *Genes* **13**, 198 (2022).

252. Piazza, A. *et al.* Cohesin regulates homology search during recombinational DNA repair. *Nat Cell Biol* **23**, 1176–1186 (2021).

253. Yang, J. H., Brandão, H. B. & Hansen, A. S. DNA double-strand break end synapsis by DNA loop extrusion. *Nat Commun* **14**, 1913 (2023).

254. Di Noia, J. M. & Neuberger, M. S. Molecular Mechanisms of Antibody Somatic Hypermutation. *Annual Review of Biochemistry* **76**, 1–22 (2007).

255. Peters, J.-M. How DNA loop extrusion mediated by cohesin enables V(D)J recombination. *Current Opinion in Cell Biology* **70**, 75–83 (2021).

256. Zhang, Y. *et al.* The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. *Nature* **573**, 600–604 (2019).

257. Srinivasan, D., Shisode, T., Shrinet, J. & Fraser, P. Chromosome organization through the cell cycle at a glance. *Journal of Cell Science* **135**, jcs244004 (2022).

258. Irgen-Gioro, S., Yoshida, S., Walling, V. & Chong, S. Fixation can change the appearance of phase separation in living cells. *eLife* **11**, e79903 (2022).

259. Bystricky, K. Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. *FEBS Lett.* **589**, 3014–3022 (2015).

260. Shaban, H. A. & Seeber, A. Monitoring the spatio-temporal organization and dynamics of the genome. *Nucleic Acids Res* **48**, 3423–3434 (2020).

261. Mullick, A. *et al.* The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC Biotechnology* **6**, 43 (2006).

262. Tasan, I. *et al.* Two-Color Imaging of Nonrepetitive Endogenous Loci in Human Cells. *ACS Synth. Biol.* **9**, 2502–2514 (2020).

263. Germier, T., Audibert, S., Kocanova, S., Lane, D. & Bystricky, K. Real-time imaging of specific genomic loci in eukaryotic cells using the ANCHOR DNA labelling system. *Methods* **142**, 16–23 (2018).

264. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**, 1911 (2018).

265. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**, 1479–1491 (2013).

266. Qin, P. *et al.* Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nature Communications* **8**, 14725 (2017).

267. Ma, H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* **34**, 528–530 (2016).

268. Clow, P. A. *et al.* CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci with one guide RNA per locus. *Nat Commun* **13**, 1871 (2022).

269. Chen, B., Zou, W., Xu, H., Liang, Y. & Huang, B. Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag. *Nature Communications* **9**, 5065 (2018).

270. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *PNAS* **112**, 3002–3007 (2015).

271. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).

272. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

273. Yang, H. *et al.* Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *International Journal of Molecular Sciences* **21**, 6461 (2020).

274. Zhang, J.-P. *et al.* Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol* **18**, (2017).

275. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).

276. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).

277. Yu, Y. *et al.* An efficient gene knock-in strategy using 5′-modified double-stranded DNA donors with short homology arms. *Nature Chemical Biology* **16**, 387–390 (2020).

278. Canaj, H. *et al.* Deep profiling reveals substantial heterogeneity of integration outcomes in CRISPR knock-in experiments. *bioRxiv* 841098 (2019) doi:10.1101/841098.

279. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids* **4**, e264 (2015).

280. Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* **7**, 540–546 (2006).

281. Sun, L., Fu, X., Ma, G. & Hutchins, A. P. Chromatin and Epigenetic Rearrangements in Embryonic Stem Cell Fate Transitions. *Frontiers in Cell and Developmental Biology* **9**, (2021).

282. Mattout, A. & Meshorer, E. Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Curr Opin Cell Biol* **22**, 334–341 (2010).

283. Herbert, S. *et al.* Chromatin stiffening underlies enhanced locus mobility after DNA damage in budding yeast. *EMBO J.* **36**, 2595–2608 (2017).

284. Backlund, M. P., Joyner, R., Weis, K. & Moerner, W. E. Correlations of three-dimensional motion of chromosomal loci in yeast revealed by the double-helix point spread function microscope. *Mol Biol Cell* **25**, 3619–3629 (2014).

285. Shukron, O., Hauer, M. & Holcman, D. Two loci single particle trajectories analysis: constructing a first passage time statistics of local chromatin exploration. *Scientific Reports* **7**, 10346 (2017).

286. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641-1659.e26 (2020).

287. Dekker, J. & Misteli, T. Long-Range Chromatin Interactions. *Cold Spring Harb Perspect Biol* **7**, a019356 (2015).

288. Masui, O. *et al.* Live-Cell Chromosome Dynamics and Outcome of X Chromosome Pairing Events during ES Cell Differentiation. *Cell* **145**, 447–458 (2011).

289. Shao, S. *et al.* Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Res* **44**, e86 (2016).

290. Shy, B. R., MacDougall, M. S., Clarke, R. & Merrill, B. J. Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells. *Nucleic Acids Research* **44**, 7997–8010 (2016).

291. Sabate, T., Zimmer, C. & Bertrand, E. Versatile CRISPR-Based Method for Site-Specific Insertion of Repeat Arrays to Visualize Chromatin Loci in Living Cells. *Methods Mol Biol* **2532**, 275–290 (2022).

292. Germier, T. *et al.* Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophysical Journal* **113**, 1383–1394 (2017).

293. Schwartz, F. *et al.* A dominant positive and negative selectable gene for use in mammalian cells. *Proc Natl Acad Sci U S A* **88**, 10416–10420 (1991).

294. Grimm, J. B. *et al.* A General Method to Improve Fluorophores Using Deuterated Auxochromes. *JACS Au* **1**, 690–696 (2021).

295. Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nature Communications* **7**, 11046 (2016).

296. Feng, S. *et al.* Improved split fluorescent proteins for endogenous protein labeling. *Nature Communications* **8**, 370 (2017).

297. Lambert, G. G. *et al.* Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing. *PLOS Biology* **18**, e3000936 (2020).

298. Hirano, M. *et al.* A highly photostable and bright green fluorescent protein. *Nat Biotechnol* **40**, 1132–1142 (2022).

299. Ghosh, R. P. *et al.* A fluorogenic array for temporally unlimited single-molecule tracking. *Nature Chemical Biology* **15**, 401–409 (2019).

300. Gustavsson, A.-K., Ghosh, R. P., Petrov, P. N., Liphardt, J. T. & Moerner, W. E. Fast and parallel nanoscale 3D tracking of heterogeneous mammalian chromatin dynamics. *MBoC* **33**, ar47 (2022).

301. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).

302. Hou, Y. *et al.* Optogenetic Control of Background Fluorescence Reduction for CRISPR-Based Genome Imaging. *Anal. Chem.* **94**, 8724–8731 (2022).

303. de Wit, E. & Nora, E. P. New insights into genome folding by loop extrusion from inducible degron technologies. *Nat Rev Genet* 1–13 (2022) doi:10.1038/s41576-022-00530-4.

304. Müller, M. *et al.* The role of pluripotency factors to drive stemness in gastrointestinal cancer. *Stem Cell Research* **16**, 349–357 (2016).

305. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

306. Ren, B. *et al.* High-resolution Hi-C maps highlight multiscale 3D epigenome reprogramming during pancreatic cancer metastasis. *Journal of Hematology & Oncology* **14**, 120 (2021).

307. Díaz, N. *et al.* Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. *Nat Commun* **9**, 4938 (2018).

308. Brown, J. M., De Ornellas, S., Parisi, E., Schermelleh, L. & Buckle, V. J. RASER-FISH: non-denaturing fluorescence in situ hybridization for preservation of three-dimensional interphase chromatin structure. *Nat Protoc* 1–26 (2022) doi:10.1038/s41596-022-00685-8.

309. Meshorer, E. *et al.* Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. *Developmental Cell* **10**, 105–116 (2006).

310. Fangman, W. L. & Brewer, B. J. A question of time: Replication origins of eukaryotic chromosomes. *Cell* **71**, 363–366 (1992).

311. Zhang, H. *et al.* CTCF and transcription influence chromatin structure re-configuration after mitosis. *Nat Commun* **12**, 5157 (2021).

312. Zhang, H. *et al.* Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* 1–5 (2019) doi:10.1038/s41586-019-1778-y.

313. Zhang, H. & Blobel, G. A. Genome folding dynamics during the M-to-G1-phase transition. *Current Opinion in Genetics & Development* **80**, 102036 (2023).

314. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183–1186 (2002).

315. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451–464 (2005).

316. Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. *Annual Review of Cell and Developmental Biology* **35**, 357–379 (2019).

317. Brandão, H. B., Gabriele, M. & Hansen, A. S. Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Current Opinion in Cell Biology* **70**, 18–26 (2021).

318. Charoensawan, V., Martinho, C. & Wigge, P. A. "Hit-and-run": Transcription factors get caught in the act. *BioEssays* **37**, 748–754 (2015).

319. Cramer, P. Organization and regulation of gene transcription. *Nature* 1–10 (2019) doi:10.1038/s41586-019-1517-4.

320. Bertrand, E. *et al.* Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell* **2**, 437–445 (1998).

321. Chen, H. *et al.* Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).

322. Pradhan, B. *et al.* The Smc5/6 complex is a DNA loop-extruding motor. *Nature* 1–6 (2023) doi:10.1038/s41586-023-05963-3.

323. Iacovoni, J. S. *et al.* High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *EMBO J* **29**, 1446–1457 (2010).

Table of figures

Résumé :

L'étude de l'organisation 3D du génome a révélé l'existence de boucles de chromatine et des Topologically Associating Domains (TADs) de l'ordre de plusieurs centaines de kilobases, créés par l'anneau de cohésine par le processus d'extrusion de boucle d'ADN. Cependant, ces structures ont été caractérisées presque exclusivement par des techniques de génomique et d'imagerie de cellules fixées, leur dynamique temporelle reste donc peu comprise. Par exemple, la durée des contacts créés par extrusion de boucles n'est pas définie et des paramètres majeurs de ce processus comme la durée de vie des contacts ancre-ancre et la vitesse d'extrusion in vivo sont toujours inconnus.

Pour répondre à cette lacune, j'ai quantifié la dynamique de l'extrusion de boucle cohésine-dépendante en visualisant et suivant dans le temps plusieurs paires d'ancres de boucles dans des cellules humaines vivantes.

Il est attendu que l'extrusion de boucle soit identifiée par une diminution progressive de la distance ancre-ancre. Cependant, cette signature est occultée par la dynamique stochastique de la chromatine, les ancres de boucles pouvant entrer en contact même sans extrusion. De plus, mesurer la distance ancre-ancre à partir d'images fluorescentes est rendu difficile par plusieurs sources d'erreurs comme les erreurs aléatoires liées à la localisation de points fluorescents.

Pour estimer les conditions expérimentales qui permettent de détecter et quantifier l'extrusion de boucles malgré ces complications, j'ai utilisé des simulations de polymères et modélisé le processus d'extrusion de boucle in silico. De plus, j'ai testé et validé de nouvelles méthodes d'analyse pour quantifier les boucles de chromatine à partir d'images statiques (e. g. à partir d'images d'ancres de boucles acquises par DNA FISH), estimer la fraction, fréquence et durée de vie des contacts ancre-ancre, ainsi qu'estimer la vitesse d'extrusion effective in vivo à partir d'images dynamiques.

En se basant sur les résultats des simulations de polymères, j'ai taggué par fluorescence de multiples ancres de boucles et TADs dans des cellules vivantes par le système CRISPR/Cas9. Nous avons conclu que les contacts entre les ancres étaient peu fréquents et de courte durée, par rapport à la durée du cycle cellulaire. Cependant, les boucles sont presque constamment soumises à l'extrusion par la cohésine. En comparant les résultats de modélisation et les expériences, nous avons pu estimer des paramètres biophysiques généraux de la dynamique d'extrusion de boucles.

Ces résultats suggèrent que l'extrusion de boucles de chromatine cohésine-dépendante est un processus hautement dynamique, qui crée des interactions à longue portée transitoires plutôt que des contacts stables. Mes résultats aideront à comprendre quantitativement des processus biologiques fondamentaux qui utilisent les contacts transitoires mais à longue distance créés par l'extrusion de boucles, comme la réparation de l'ADN et la régulation de l'expression des gènes.

Mots clés : [Organisation 3D du génome, Extrusion de boucles de chromatine, Imagerie en cellules vivantes, Simulations de polymères, Cohésine]

Modeling, visualizing and quantifying chromatin loop extrusion dynamics in living human cells.

Abstract:

Studies of spatial genome organization have revealed the existence of chromatin loops and Topologically Associating Domains (TADs) of several hundred kilobases in size, which are created by the cohesin ring complex through a process of DNA loop extrusion. However, these structures have been characterized almost exclusively by genomic techniques and fixed cell imaging, thus their temporal dynamics are still poorly understood. For example, it is not clear whether loop extrusion creates stable or transient contacts at loop anchors and key parameters of this process, including loop lifetime and extrusion speed, remain unknown.

To address this gap, my thesis aims to quantify the dynamics of cohesin-dependent loop extrusion by visualizing and tracking in time pairs of anchors at several loops and TADs in living human cells.

Extrusion is expected to manifest itself as a progressive decrease in anchor-anchor distances. However, this signature is obscured by stochastic motions of the chromatin, whereby anchors can occasionally come into contact even without extrusion. Furthermore, measuring the anchor-anchor distance from fluorescent images is complicated by several sources of uncertainties, such as unavoidable random errors in the computational localization of fluorescent spots.

To evaluate the experimental conditions under which one can expect to detect and quantify loop extrusion despite such complications, I first performed an analysis in silico using polymer simulations that account for loop extrusion. Using these simulations, I also tested and validated novel analysis methods to quantify chromatin loop dynamics from static imaging (e. g. from DNA-FISH images of loop anchors), and to estimate the lifetime and frequency of anchor contacts, as well as the effective loop extrusion speed from dynamic imaging in vivo.

Using the simulation results as guidelines, we fluorescently labelled multiple loop and TAD anchors in human cells using the CRISPR/Cas9 system and tracked the loop anchors by live-cell imaging. Based on our analysis of the imaging data, we found that contacts between the two loop anchors are infrequent and short-lived as compared to the cell cycle duration. However, loops were found to be almost constantly extruded by cohesin. By comparing simulations and experimental data, we could estimate key biophysical parameters of loop extrusion dynamics including loop lifetimes and extrusion speed.

Our results suggest that cohesin-dependent loop extrusion is a highly dynamic process, which creates transient long-range interactions rather than stable contact s. Our findings will help to quantitatively understand biological processes that involve short-lived but long-range contacts created by loop extrusion, including mechanisms of DNA repair and gene regulation.

Keywords: [3D genome organization, Chromatin loop extrusion, Live-cell imaging, Polymer simulations, Dynamics, Cohesin]