# Imaging the dynamic architecture of chromatin at the single cell level

Laura Caccianini

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie

# Imagerie de l'architecture dynamique de la chromatine dans la cellule unique.
# Probing the interplay between CTCF and Cohesin with single molecule imaging in living cells.


Soutenue par

**Laura Caccianini**

Le 09/09/2019


École doctorale n°564

**Physique en Ile-de-France**

Spécialité

**Physique**

Composition du jury :

| | |
|---|---|
| Marco COSENTINO-LAGOMARSINO<br>DR, IFOM, Italy | *Président du jury* |
| Christof GEBHARDT<br>PU, Universitat Ulm, Ulm, Germany | *Rapporteur* |
| Marcelo NOLLMAN<br>DR, CBS, Montpellier, CNRS, France | *Rapporteur* |
| Kerstin BYSTRICKY<br>PU, Paul Sabatier, CBI, Toulouse, France | *Examinateur* |
| Martial MARBOUTY<br>CR, Institut Pasteur, CNRS, France | *Examinateur* |
| Natalie DOSTATNI<br>PU, Institut Curie, SU, France | *Examinateur* |
| Mathieu COPPEY<br>CR, Institut Curie, CNRS, France | *Directeur de thèse* |
| Maxime DAHAN†<br>DR, Institut Curie, CNRS, France | *Directeur de thèse* |

PSL UNIVERSITÉ PARIS

institutCurie

*À Maxime*

*De tout, il resta trois choses:*
*la certitude que tout était*
*en train de commencer,*
*la certitude qu'il fallait continuer,*
*la certitude que cela serait interrompu avant que d' être terminé.*
*Faire de l'interruption, un nouveau chemin,*
*faire de la chute, un pas de danse,*
*faire de la peur, un escalier,*
*du rêve, un pont,*
*de la recherche...*
*une rencontre.*

*Fernando Pessoa*

# Contents

# Introduction

The nucleus is a very complex and crowded environment. It is the centre where important cell processes are regulated. This crowded environment however is highly organised. The complexity of this system made it a biologist's nightmare and a physicist's playground. During my thesis I tackled various challenging questions related to mechanisms that take place in the nucleus, mainly with single molecule imaging techniques. Hereafter I introduce some of the projects I was involved in and expand more on the main subject of my work that is described in this thesis.

I first approached the mammalian nucleus during my internship, when I started studying the target search process of DNA-binding proteins, such as transcription factors. When searching for their target, transcription factors interact non specifically with various DNA sequences. To better understand this process I performed single molecule tracking experiments on transcription factors-like proteins. To mimic transcription factors I worked with TALE (Transcription Activator Like Effectors) proteins, which can be designed to target any sequence of interest by changing two central residues (the 12th and 13th) in the DNA binding domain. Thanks to the versatility of this method, 6 TALEs targeting sequences of different lengths were designed and studied. In particular, we characterised proteins with cognate sequences of 6bp (basepairs), 10bp, 14 bp, 18 bp, 22bp and 26 bp. Non-specific interactions, revealed by single molecule imaging and tracking, are not dramatically affected by the length of the DNA binding domain of the protein, except for the shortest one (6bp and 10bp) that deviates from the other samples. The distributions of residence times are always characterised by a long tail (power law-like) indicating that the non-specific interaction stability does not depend on the length of the DNA binding domain. Our intrerpatation of such observation is that while short proteins experience non-specific interactions because of the high redundancy for short sequences in the genome, longer proteins are more tolerant to mismatches and can thus be kidnapped by quasi-cognate sequences (as binding a few nucleotides is energetically suf-

ficient to perform a transient binding). This subject will not be treated in this manuscript as it would require a long and detailed discussion. We are currently concluding this work and hopefully our results will soon be published. The study of target search in the nucleus of mammalian cells raises the question of chromatin accessibility and that is how I became interested in chromatin spatial organisation and dynamics.

Moved by the curiosity on nuclear organisation, I started tracking genomic loci, mainly in mouse embryonic stem cells. We started a nourishing collaboration with Prof Edith Heard and her team that brought me to study the dynamic behaviour of loci at known genomic distances and with different transcriptional states (active/silenced) on the X chromosome. The preliminary results are promising but, as this project is still ongoing, further analysis are still needed before any conclusion can be made.

This manuscript recapitulates the results of the main project run during my PhD: a single molecule study of the interplay between two nuclear factors that have been found to play a crucial role in the regulation of chromatin architecture, CTCF (CCCTC binding factor) and Cohesin. This topic will be treated in detail in the chapters of this manuscript. The thesis is divided into 4 chapters; henceforth I describe the content of each.

Chapter 1 is dedicated to introducing the tight link between structure and function. After a description of the advantages arising from three-dimensional, dynamic, folding of the genome, I describe the main techniques used to capture chromatin structure. A brief overview is then given on the multi-scale organisation of DNA with a focus on Topologically Associating Domains (TADs), which constitute the scale of interest of this study. The following paragraph is devoted to describe an hypothesis of mechanism for TAD formation that is gaining more and more credit: the loop extrusion hypothesis. The chapter is then concluded with two paragraphs on the protagonists of the loop extrusion model and object of this study, the CCTC binding factor (CTCF) and Cohesin.
Chapter 2 describes the techniques deployed for the goal of this work. In particular a brief introduction on optical microscopy is presented, followed by a paragraph on single molecule localisation microscopy. A detailed description of the cell lines and cell culture condition is also provided. The last section of the chapter is dedicated to a detailed description of the analysis of single molecule tracking data.
In Chapter 3 the results of single molecule tracking of the different proteins, in the context of various alterations, are presented. The first section is a short

insight on tracking of a genomic locus that has been used as a reference for the template's (chromatin) dynamics. This section is followed by the results of tracking of CTCF and by the results for Cohesin in presence or absence of CTCF, in absence of Sororin and in absence of Nipbl (two factors involved in the regulation of Cohesin binding kinetics that will be presented in detail in their respective section). The manuscript is concluded by a section where all the results are discussed and put in the context of the current literature and state of the art knowledge on the subject.

# Chapter 1

# On Chromatin Architecture

## 1.1 Chromatin structure and function

Chromatin architecture plays a crucial role throughout the whole life of a cell.
A simplistic but meaningful way to understand the importance of the relationship between structure and function is represented in figure 1.1. In this schematic picture, the blue puzzle pieces represent regulatory elements that need to get in contact to accomplish a function and the red arrows represent the idea that there is a preferential interaction, that the two regulatory elements are not brought together by just thermal motion.

Two elements that are far apart on the uni-dimensional fiber can be found in spacial proximity in a three-dimensional space. In general, jumping from 1D to a 1+n space (where $n \geq 1$) allows many more structural configurations. For a more detailed description of the implications, and the mechanisms, of the 3D chromatin organisation see (Dekker and Mirny, 2016).

This simple and strong argument supports plenty of findings that relate structure and function. Gene expression can be triggered via the regulation of chromatin structure. This can happen in mainly two ways: by chromatin compaction, and by bridging regulatory elements. These two scenarios are sketched in fig 1.1 (bridging distal regulatory elements) and in fig 1.2 (compaction regulates accessibility).

There is in fact a distinction between compact and less-compact chromatin, named respectively *heterochromatin* and *euchromatin*. The discovery of the difference in compaction (thus staining) of some chromosomal regions, or *heteropycnosis* as scientists would have said at the time, dates back to the early 20th century (Passarge)(see fig 1.3). Already in his *Heterochro-*

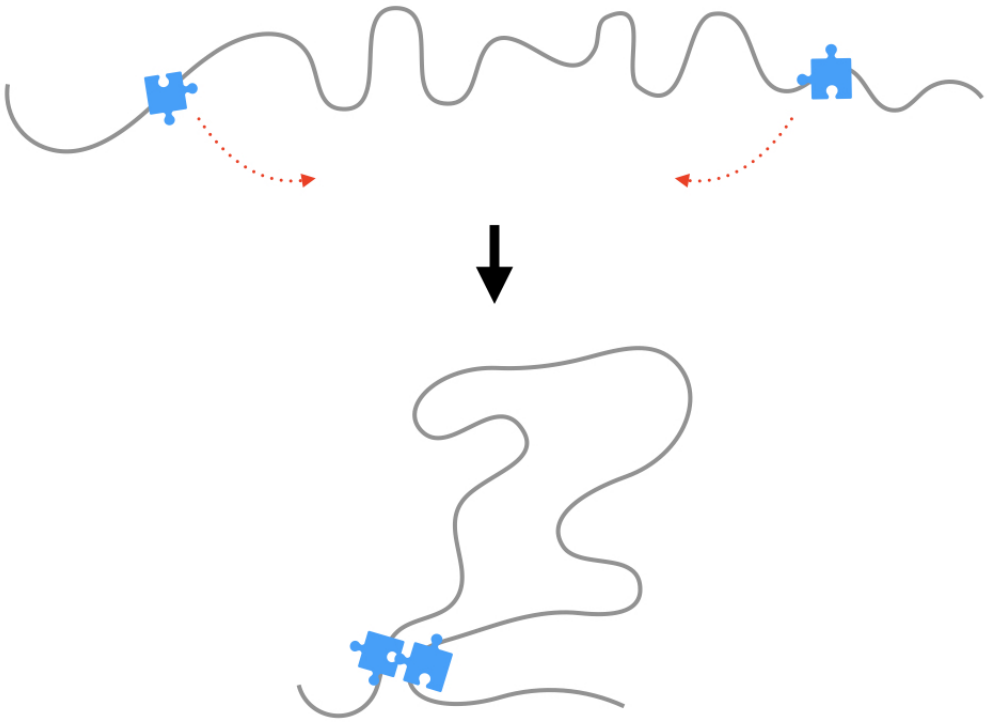


***Figure 1.1: How to promote spatial proximity of two elements on a 1D fiber.*** *A naive picture.*

*matin, Chromocentren, Chromomeren. Ber Botan Ges. 47:274-284, 1929*, Emil Heitz claims that the less stained *euchromatin* is genetically more important than *heterochromatin*. It is now known that euchromatic regions are rich in genes and the transcriptionally active ones. Very interesting works have been done also on the correlation between the state of chromatin (eu- or hetero-chromatic) and their positioning in the nucleus, but for the sake of space I will not tackle this exciting topic. More information on the subject can be found in the work of Dr Solovei (Solovei et al., 2016) and Dr. Van Steensel (van Steensel and Belmont, 2017).

Another example of structure and function relationship can be found in the different DNA repair mechanisms. In case of DNA strand break chromatin undergoes structural remodelling to facilitate the access of the repair machinery (Stadler and Richly, 2017). Gene expression and DNA repair mainly take place during interphase, but chromatin structure regulation is dramatically important for cell division. DNA shall be replicated, condensed

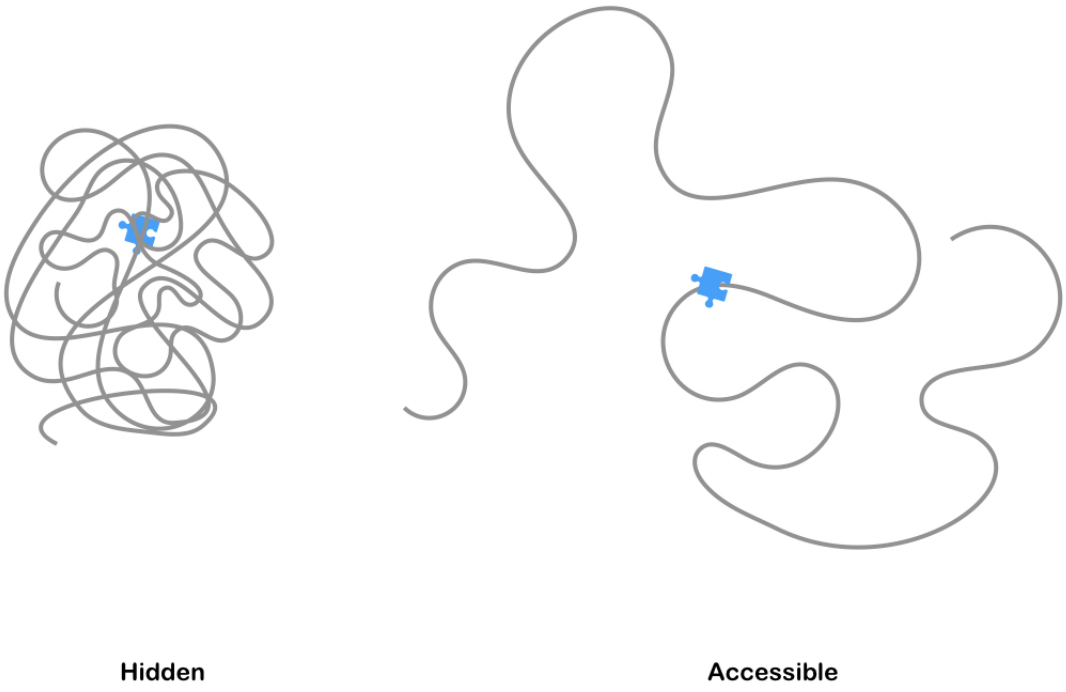



***Figure 1.2: Accessibility as a consequence of compaction.*** *Modulating the fiber compactness means modulating some targets accessibility. The blue puzzle piece could represent a promoter, a gene, a DNA damage site, ...*

and equally partitioned at the two poles of the dividing cell. Proper condensation and segregation are crucial (Gibcus et al., 2018).
Transcriptional activity also affects positioning in the nucleus. Gene rich chromosomal regions (or euchromatin) tend to be found at the center of the nucleus in most species, while heterochromatic regions are usually positioned at the periphery (Bickmore, 2013). (An interesting counterexample of "inverted nuclei" in rod photoreceptor (Falk et al., 2018)).
These are some of the demonstrations of the tight relation between DNA structure and function. The examples named above actually concern different length-scales and are regulated by different mechanisms. While DNA repair processes are more related to epigenetic marks, thus less than kilobasepairs (kbp) length-scale, gene expression regulation happens at the scale that spans from $100_s$ of kbp up to 1 Mbp. In the following sections I will describe more in detail the classification of chromatin structures, or domains,

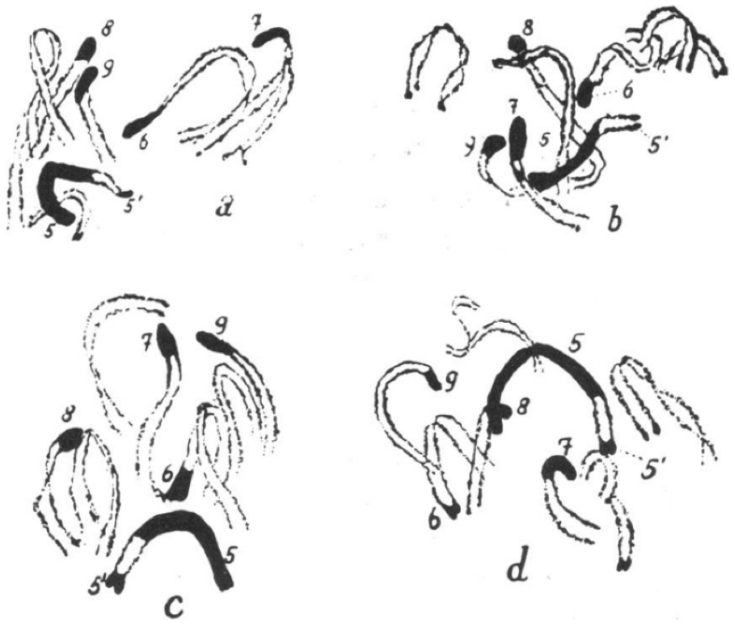



***Figure 1.3: The first picture of heterochromatin and euchromatin.*** *Darkly stained heterochromatin and lightly stained euchromatin in Pelliaepiphylla (from textitDas Heterochromatin der Moose.I.Jahrb Wiss Bot 69:762-818,1928). Adapted from (Passarge)*

in relation with their characteristic length-scale.

### 1.1.1 Chromatin structure: how to characterize it

In the last decade, the investigation of chromatin structure has been pushed forward by the establishment of different techniques. Overall, chromosome conformation capture techniques ($3C, 4C, ..., Hi-C$) really made a difference (Dekker, 2002), (Lieberman-Aiden et al., 2009). The $3C$ based techniques are a quantitative method to characterise the interaction between genomic loci.

Conformation capture techniques are based on the following steps (graphically resumed in fig 1.4):

1. cell fixation (crosslinking);

2. chromatin digestion;

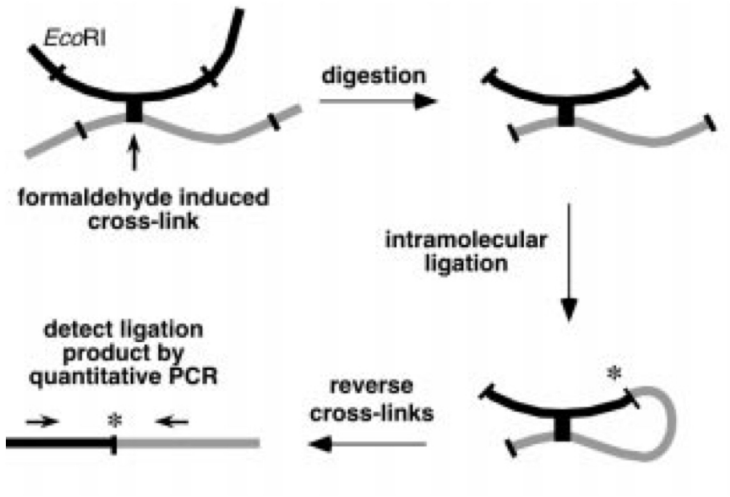



***Figure 1.4: Steps of a conformation capture experiment.*** *Sequential steps of a chromosome conformation capture experiment. The protocol here represented is adapted from the first paper on the technique: (Dekker, 2002). Further implementation of the method (like 5C or Hi-C) have pushed forward the sequencing depth and the genomic resolution but are based on the same principles shown here.*

3. intramolecular ligation;
4. reverse crosslinking/precipitation of the segments;
5. quantitative PCR or sequencing.

The output of a conformation capture experiment is an ensemble averaged matrix of interaction, which is nothing but an histogram of contacts. Chromatin and its complex structure tickled the curiosity of scientists from different domains. Not only biologists, bioinformaticians and polymer physicist, but also scientists from the microscopy community are actively contributing to the characterization of chromatin architecture. Historically, electron microscopy played a crucial role in revealing chromatin structure; images of compartmentalisation were produced in the 1920*s* (as mentioned above (Passarge) and fig 1.3), various nuclear bodies were discovered back in the late 1960*s* (Monneron and Bernhard, 1969) and a very recent work disclaimed the existence of the scholar "30 nm fibre" by revealing filaments with a diameter that spans from 5 to 24 nm, and whose packing varies depending on cell

cycle (Ou et al., 2017). Electron microscopy is a very powerful techniques as it reaches the highest resolution amongst imaging techniques (atomic length-scales), unfortunately it does not allow (yet?) for a direct identification of the object of interest (via a specific staining or a tag, for example). That is where optical microscopy fills the gap. Optical microscopy resolution is lower but we can pinpoint the desired entity (a structure, a protein, a complex) with a specific fluorescent probe. Fluorescence In Situ Hybridization (FISH) (Langer-Safer et al., 1982) has been the most common approach to visualise genomic loci and investigate chromatin spatial organisation; just to cite two examples: the work of the Cremer brothers where they showed for the first time the chromosomal territories (Cremer and Cremer, 2001), and another study showing that silenced genes on the inactivated X chromosome in differentiated cells are subjected to physical compaction (Chaumeil, 2006).

In less than a decade huge advancements have been made in the field of chromatin imaging. A considerable improvement comes from the development of labelling strategies like Oligopaint probes (Beliveau et al., 2012); inspired from the "classical" immunostaining (i.e. FISH), DNA Oligopaint is based on a precise design of probes (PCR-based) that incorporate a fluorophore. The renewability of the probes enables multi-color imaging and importantly it makes the tool compatible with high-sampling localisation techniques [1] like STORM (Rust et al., 2006), or dSTORM (I will use "STORM" in both cases). Thanks to Oligopaint it is possible to label specific sequences and to image them with a localisation precision of less than 20 nm. In his Oligopaint-STORM based work in Drosophila, Boettinger showed the correlation between domain extension and epigenetic state, confirming that silenced domains are more condensed (Boettiger et al., 2016). Shortly after, these observations have been confirmed by study that combined STORM and another technique known as Structured Illumination Microscopy (SIM)[2]; in his paper, Cattoni and colleagues showed TAD in Drosophila in the whole nucleus (Cattoni et al., 2017). Similar studies have been performed in mammalian cells (Bintu et al., 2018) which nicely shows the heterogeneity of a cell population, and some eventually achieved a massive, consecutive, labelling of 8 Mbp of chromosome 19 (Nir et al., 2018). The insights of super-resolution imaging works on mammalian TADs will be discussed in 1.2. Importantly this works revealed a significant cell-to-cell variability, suggesting a scenario towards which more and more evidences point: that domains like TADs and

[1] with "high-sampling localisation" I mean the super resolution techniques that are based on numerous individual localisations.

[2] SIM resolution improves canonical wide field, diffraction limited, imaging of a factor 2 while STORM can push the improvement to a factor 10.

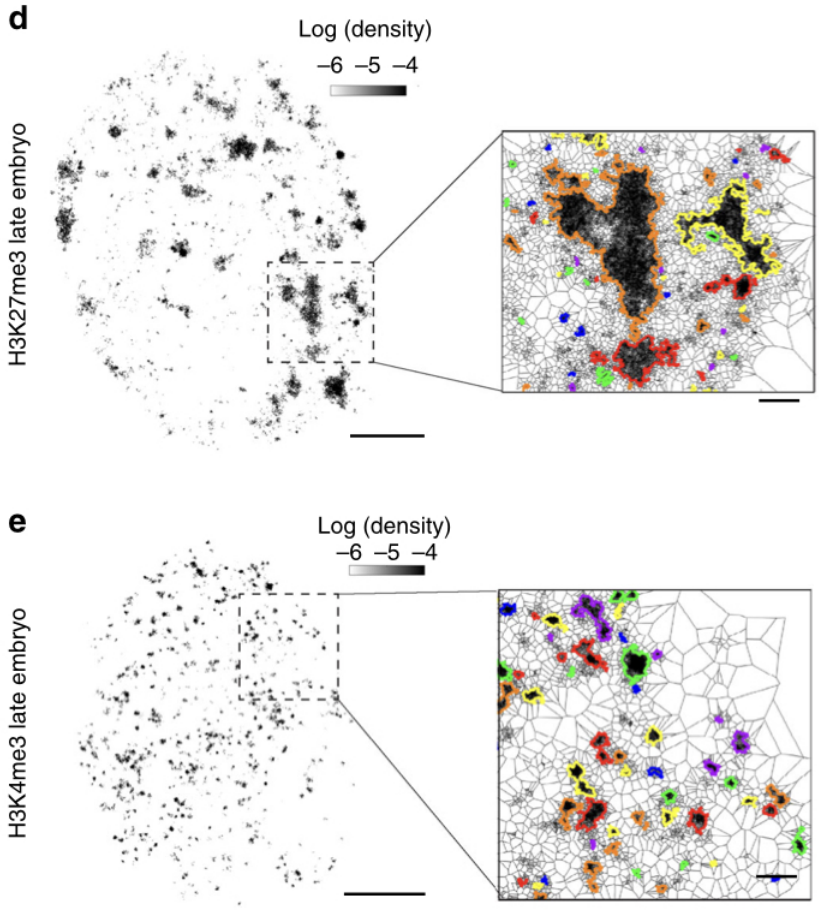



***Figure 1.5: Super resolved images of domains in Drosophila.*** *STORM image of transcriptionally repressed (top, panel d) and active (bottom, panel e) domains. Adapted from (Cattoni et al., 2017).*

compartments are generated by low frequencies, yet specific, interactions triggered by transcriptional and epigenetic state. Examples of single cell super-resolved images of TADs in Drosophila are reported in fig 1.5.

Cell-to-cell variability is actually a natural symptom of the mechanism that is thought to be behind loop formation. According to the loop extrusion hypothesis, loops are not fixed structures. This makes them difficult to directly visualise with both conformation capture methods and microscopy conventional techniques.

Moreover, combining results from conformation capture experiments and FISH, or FISH inspired techniques, is not always trivial. Despite the two ap-

proaches are often used to cross validate their respective results, sometimes the findings are not necessarily coherent. This has to do with the intrinsic features of the techniques: while 3C is capable of detecting interactions at the molecular scale (sub-nanometric), FISH is limited by the diffraction limit (hundreds of nanometers see section 2.1), while super-resolution methods can reach the tens of nanometer scale (see fig 1.6). Furthermore capture techniques are usually performed on tens of thousands of cells, while imaging is performed on tens of cells. When tackling the same question with different methods we should be aware of what can be resolved and welcome the discrepancies between the results to better understand the underlying structures, as pointed out in (Dekker, 2016).

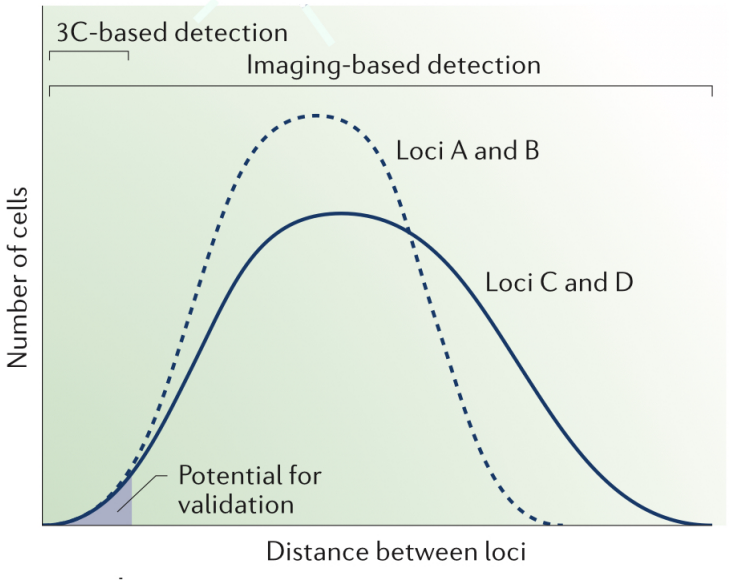



***Figure 1.6: How to compare apples and oranges: the parameter space overlap of 3C and FISH experiments.*** *Conformation capture and imaging methods do not cover the same scale of physical interaction, or more in general spacial proximity. Discrepancies could be an alarm sign of artefacts, but also an indicator of more complex mechanisms and structures. Adapted from (Dekker, 2016).*

Fudenberg and Imakaev tried to build bridges between the results of the 'C' techniques (conformation capture techniques) and FISH. In their publication (Fudenberg et al., 2017), they point out that while FISH can report the cell-to-cell variability and is free from binning issues differently from Hi-C,

it can hardly validate Hi-C results as it is a very low throughput techniques. Consequently striking perturbations in the Hi-C phenotype are difficult to validate by FISH as rare events are not well represented in the experimental probability distribution of distances. An example of both matching and not-matching FISH and Hi-C results for two loci is reported in fig 1.7. These limitations to imaging techniques still stand for the cutting edge super-resolution approaches cited above. However, the recent advancements in the labelling approaches (i.e. DNA-OligoPAINT and CRISPR-Cas9 based techniques) significantly improved the resolution and therefore the localisation precision, which allows better identification of looping. In these two different works, Fudenberg and Imakaev showed that conformation capture data should be treated carefully as contact probabilities cannot be simply translated into spatial distances distributions (Imakaev et al., 2015). In (Fudenberg et al., 2017) they show how the hypothesis of dynamic looping in polymer simulations can reconcile potentially Hi-C and FISH non-matching results. Fostering dynamic exchanges between different experimental approaches seem to be the best strategy to investigate chromatin architecture.

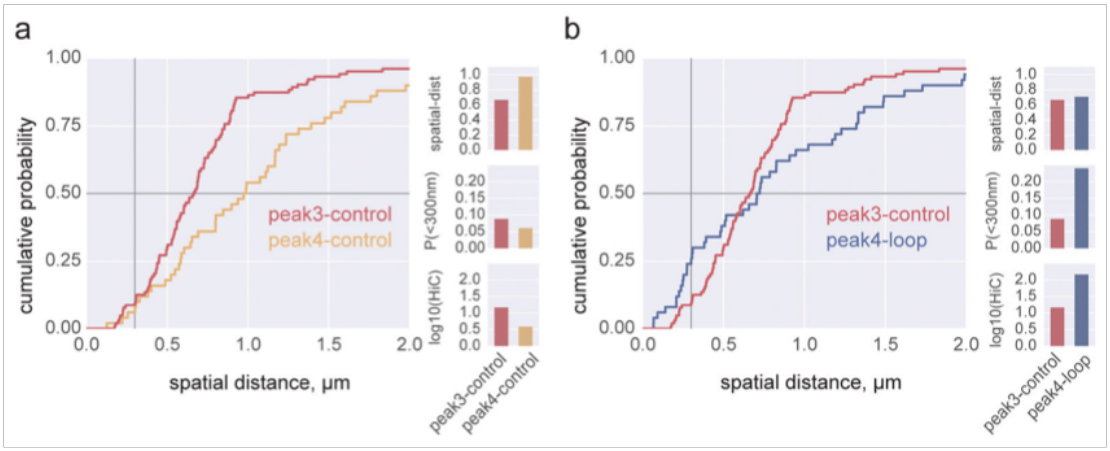



***Figure 1.7: 2-loci distance probed with FISH and Hi-C. a*** *Two loci shows increased spatial distance in FISH but decreased contact frequency in Hi-C.* ***b*** *Pair of loci showing increased spatial distance but decreased Hi-C counts. Adapted from (Fudenberg et al., 2017), data from (Rao et al., 2014).*

The techniques presented above, both on the biochemistry and imaging side, are based on cell fixation, hence they cannot provide information on the dynamics of chromatin folding. Nevertheless, tracking fluorescent genomic loci in living cells can complement this static picture.

**Table 1**
DNA labeling tools for use in eukaryotic living cells.

| System | Origin | Applications, advantages, drawbacks | Genome editing | References |
|---|---|---|---|---|
| FROS (fluorescent repressor operator system) | *E. coli* chromosome (lactose); Tn10 transposon (tetracycline gene); phage lambda | Yeast, difficult in metazoans; replication blockage (fragile sites); potential chromatin disruption, interference with transcription | Required; insertions 5-10 kb of arrays of repetitive DNA sequences | [14,23,22] |
| TALE (transcription activator like effector proteins) | *Xanthomonas oryzae* transcription activator | Any cell type; restricted to repetitive sequences | No | [17,18] |
| CRISPR/Cas9 | Streptococcus pyrogenes (Sp), *Neisseria meningitidis* (Nm), and *Streptococcus thermophilus* (St1) | Any cell type; restricted to repetitive sequences or numerous sites (several kb); local DNA unwinding/triple helix formation | No | [15,16] |
| Suntag | Antibody | Potentially any cell type; very specific, lack of versatility; final complex very large (up to 1400kDa); high probability of instability, aggregation and interference with function of protein studied | Required; insertion of protein binding sites | [19] |
| ANCHOR (ParB/INT) | Burkholderiaecae chromosome partition systems | Any cell type; versatile; negligible interference with DNA processing, transcription | Required; insertions of 0,4-1 kb unique sequences | [20] |
| Fluorescent dNTPs | | Bulk labeling; whole genome; unspecific; labeling requires single cell injections or 'rubbing' | No | [103] |
| Histones (H2B-GFP etc.) | | Bulk labeling; whole genome; unspecific; stable fluorescence; photoactivatable fusion allows activating individual chromosomes | No | [49,93] |

***Figure 1.8: An exhaustive summary of the labelling option of live imaging of chromosomal loci.*** *Adapted from (Bystricky, 2015)*

.

There are different options for labelling genomic loci. The classical approach is based on fluorescence repressor operator systems (i.e. the LacO approach), or FROS as called in table 1.8, adapted from (Bystricky, 2015). Such technique consists in the insertion of an array that can be targeted by a specific repressor which is typically coupled to a fluorescent probe. But the FROS approach relies on big ectopic insertions that may not reflect the dynamics of an endogenous system. From this point of view the novel ANCHOR system seems to be less invasive as it relies on a dynamic ParB-ParB and ParB-DNA interactions (Germier et al., 2018). A more spread approach is based on the CRISPR-Cas9 system which allows endogenous homologous recombination and targeting of non-repetitive sequences as shown in (Gu et al., 2018).

Live imaging of DNA loci contributes with essential parameters for the characterisation of the chromatin as a polymer. Inferring properties such as elasticity (is it elastic?) and the characteristic diffusion time-dependance, is crucial to build finer models for DNA dynamic folding.

### 1.1.2 Chromatin structures span over different length-scales

We know today that there are multiple levels of DNA organisation.

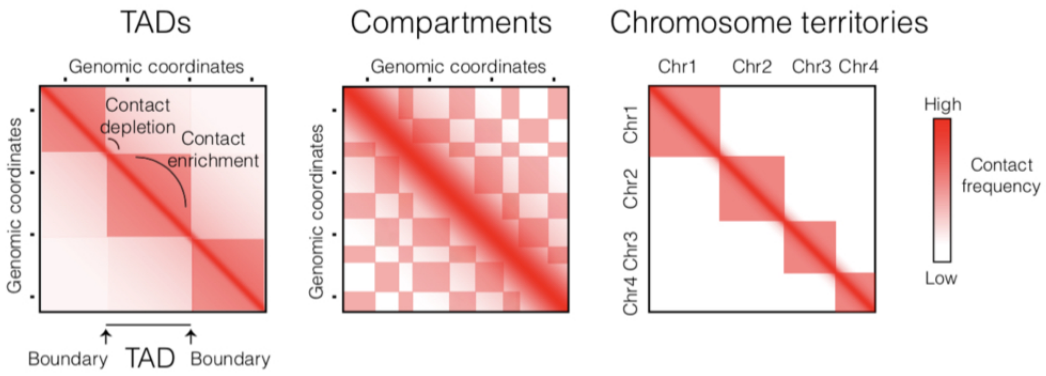



***Figure 1.9: Chromatin is packed in different domains over a wide range of length-scales.*** *A picture of the wide, multi-scale genome organisation as it appears in a chromosome conformation map with the different patterns arising from compartmentalisation, TADs and chromosome territories. The signature is domain-specific and not hierarchical. The figure, as many in the literature, is lacking the loop which in conformation capture maps typically appear as corner peaks (see (Fudenberg et al., 2016)). Adapted from (Szabo et al., 2019).*

The smallest domain of structural organisation we can think of (in eukaryotes) is the nucleosome: $\sim$ 150 bp of DNA wrap/are wrapped around histones. At the 10s to 100s kbp scale, genome is organised in the so-called Topologically Associating Domains (TADs), domains characterised by higher intra-domain interactions rather than inter-domain, for mammalian genomes. On the larger genomic scale of Mbp we find compartments which correspond to the euchromatic and heterochromatic regions cited in section 1.1. These domains can be either A compartments (gene-rich, transcriptionally active, less compact, typically positioned at the center of the nucleus) ei-

ther B compartments (gene-poor, more compact, located at the periphery of the nucleus). The largest domain identified is the chromosomal territory; chromosome do not intermingle thus each of them consist in forming its own domain.
As shown in fig 1.9, genome packaging and domains span over three order of magnitude of physical distance: from $10nm$ up to a few $\mu m$. In the majority of genome's architecture description, loops are not mentioned as a level of organisation. Any pair of loci getting in contact are forming a loop, in this sense loops are the basic ingredients of chromatin 3D structure.

TADs and compartments give rise to specific patterns in a Hi-C map, see fig 1.9. Numerous discoveries in the mechanisms behind the regulation of chromatin structure are related to the different "phenotypes" arising in these maps. The features and differences of TADs and compartments patterns in Hi-C maps make it possible to distinguish the perturbations on each level of chromatin organisation. In the last 2-3 years we learned for example that depleting factors involved in TAD formation does not necessarily have an impact on compartments (Schwarzer et al., 2017), (Nora et al., 2017), (Haarhuis et al., 2017) and (Rao et al., 2017) (For more detailed discussion see section 1.2). I will discuss the effects of altering TAD regulation and maintenance in section 1.2, here I want to underline that it is proven that TADs and compartment do not arise from the same physical process. A model has been recently proposed to explain the interplay of the different mechanisms behind TADs and compartments (Nuebler et al., 2018). An informative quantity for the description of chromatin spacial arrangement is the contact probability $P(s)$, which represent the renormalised contact frequency as a function of the genomic distance $s$. Interestingly, the contact probability distribution decays following a power law, which is an intrinsically scale-free function. A few examples are reported in fig 1.10.

For the sake of brevity, and coherently with the objects of this study, I will focus on TADs. In the following sections I will first introduce TADs and the proposed mechanisms for their regulation and I will continue by describing the actors involved in such process.

## 1.2 On Topologically Associating Domains (TADs)

I here intend to draw a basic picture of TADs features in the most general terms; this brief description is not meant to be exhaustive but to provide to the reader the founding elements of this work.
TADs are a recent discovery that has profoundly influenced the related fields

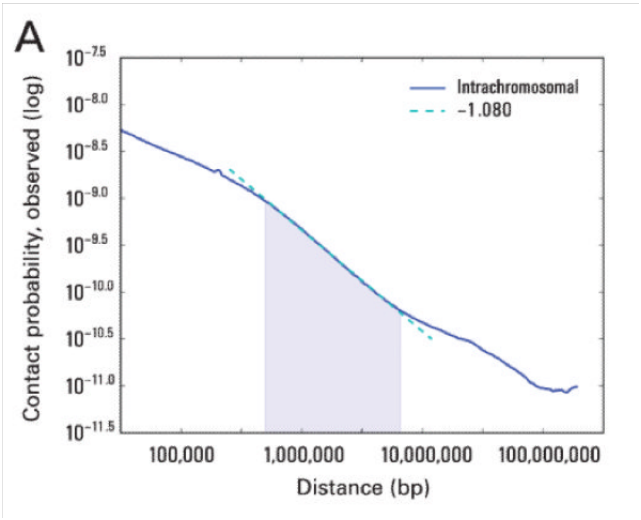



***Figure 1.10: The contact probability follows a power law decay as a function of genomic distance. From the first publication of Hi-C data.*** *Adapted from (Lieberman-Aiden et al., 2009).*

(from epigenetics to biophysics). The three seminal studies on TADs came out in 2012 (Dixon et al., 2012), (Nora et al., 2012), (Sexton et al., 2012). In figure 1.11 the reader will find a simplified timeline that marks the milestones of TADs discovery and investigation.
Topologically Associating Domains (TADs) are regions of self interacting chromatin, or regions that that tend to interact with themselves more than with other regions. TADs represent a preferential scale of functional *cis*-contacts (Zhan et al., 2017) and the disruption of this level of organisation can have dramatic consequences.

We now have more and more evidences of the In a chromosome conformation capture map (5C or Hi-C), TADs generally appear as squares with increased number of contacts, as pictured in fig 1.9.

TAD-like structures can be found in different species such as bacteria, yeasts (in Pombe but not in cerevisiae), Drosophila, and mammals, in which they were discovered. In fig 1.14 four different species are represented. On the **top left** corner there is an example of a bacterial Hi-C map. In this work (Le et al., 2013) Le and colleagues showed that in *Caulobacter* there is a series of highly self-interacting domains that they named CID (Chromosomal Interaction Domains), intercalated with super-coiled plectonemes. Despite the fact that many processes are not similar to mammalian ones, and the actors involved in the regulation of such processes are not the same of eukaryotes, bacteria present similarities with eukaryotes. It has been shown

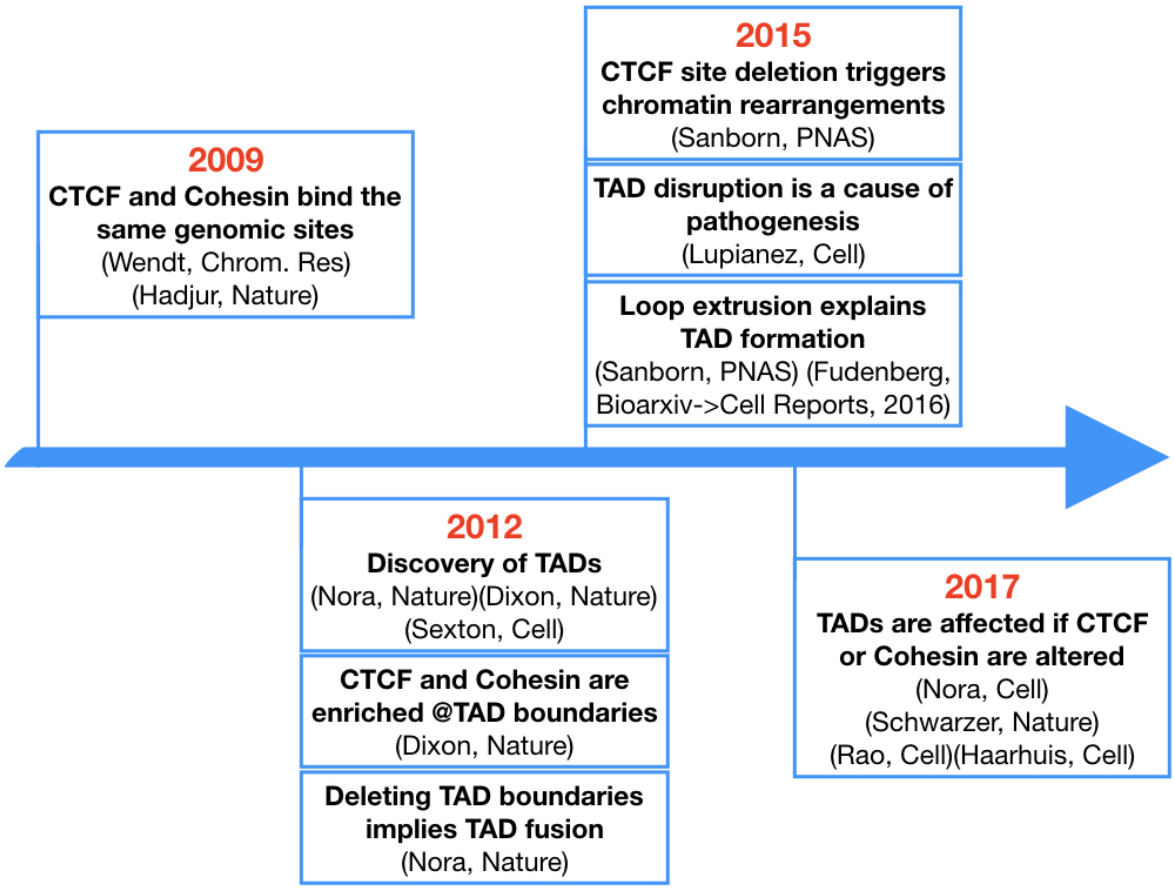



***Figure 1.11: Milestones of TADs discovery and characterisation.***

that complexes from the Structure Maintenance Complex (SMC) family (together with parS and ParB) are crucial for the establishment and maintenance of chromosomal domains (CID) (Marbouty et al., 2015). In plants (**top right**), TADs can't be easily identified, but some of the TAD's features are observed in the contact maps, yet no insulating proteins are known in plants. In *Drosophila*, TADs have been studied in whole embryos (Sexton et al., 2012); in flies these domains are classified on an epigenetic basis.
TADs have been discovered in mammals (Dixon et al., 2012), (Nora et al., 2012) and Drosophila (Sexton et al., 2012), the organisms with the most well defined structures. Mammalian TADs have very specific features: their boundaries are determined by convergent binding sites (Rao et al., 2014) of the CCCTC-binding factor (commonly known as CTCF) and an enrichment of Cohesin at this same location (Merkenschlager and Nora, 2016). They are also characterised by corner peaks. Figure 1.14 is missing yeasts, where we

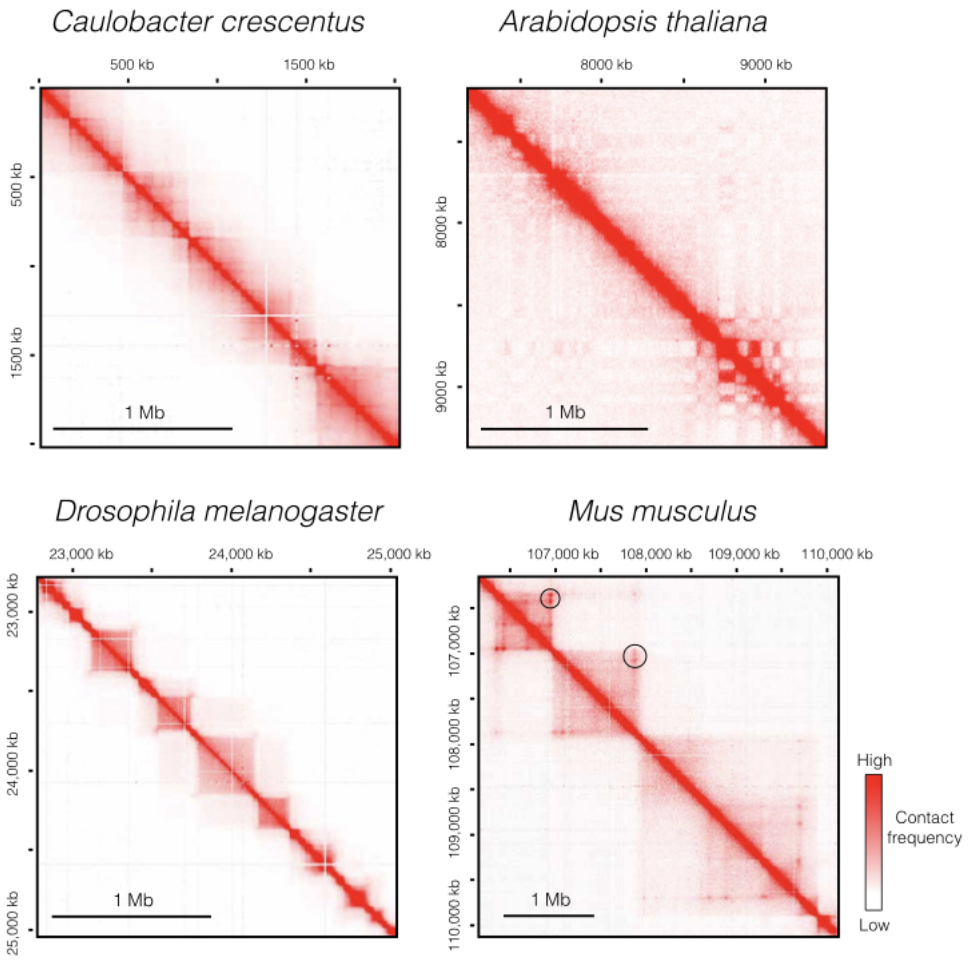



***Figure 1.12: TADs across species.*** ***Top*** *Chromosomal Interaction Domains (CIDs) are TAD-like structures in bacteria Caulobacter crescentus, and in plants, Arabidopsis thaliana, (Szabo et al., 2019).*

also find TADs (Tsochatzidou et al., 2017) [3]. TADs are also conserved in different tissues (Dixon et al., 2012) (Smith et al., 2016), and there are more

[3]In ref (Tsochatzidou et al., 2017) the authors changed the name of the domains from TADs to "insulated domains" after discussion with one of the reviewers. I find fascinating the way the lexicon is established in science. For those interested in some everyday epistemology, check the authors comment: https://computational-genomics-uoc.weebly.com/blog/tads-in-yeast-and-how-you-can-go-around-a-reviewer-if-you-are-right.

and more evidences showing that TADs are conserved throughout evolution (Krefting et al., 2018).
The current picture of domains rely on static assays, nevertheless TAD should not be thought as stable or ubiquitous structures. Single cells experiments have disclosed the high cell-to-cell variability that is smoothed in canonical Hi-C assays (averaged over tens of thousands of cells) (Nagano et al., 2013)(Flyamer et al., 2017). While compartments at tens of megabase scale are maintained in the majority of cells (the checkerboard pattern is clearly recognisable), smaller compartments are more variable, still intra-domain contacts are more conserved than inter-domain ones. Similar results have been published in microscopy-based works in Drosophila (Szabo et al., 2018) (Cattoni et al., 2017) and in mammals (Bintu et al., 2018), but despite a reasonable cell-to-cell variability TAD-like domains appear recurrent, definitely more than one would expect from stochastic interactions.
In the last 3 years, different groups have shown the important role played by different factors in the regulation and maintenance of chromatin arrangement at the TAD scale. In particular (Nora et al., 2017) have shown that CTCF is crucial to maintain TAD organisation, as depletion of CTCF leads to complete loss of TADs. Similarly Cohesin depletion has as a consequence the loss of TAD (Rao et al., 2017), as well as Nipbl degradation (Schwarzer et al., 2017). On the other hand Wapl depletion has no such dramatic effects and eventually leads to stronger TADs (Haarhuis et al., 2017). The results of these alterations are reported in fig 1.13. A key finding of these works is the demonstration of the de-coupling between TAD and compartment organisation. While TAD are heavily affected by the depletion of CTCF and Cohesin, compartments remain untouched by these perturbations. Such result is a proof of the fact that different scales of chromatin spatial organisation, TADs and compartments, are not regulated by the same mechanisms. Together, these results have significantly contributed to individuate the role played by each of these factors, which will be described more in detail in the following sections.

To complement ensemble Hi-C experiments, different groups are now dedicating their efforts to single cell Hi-C and single cell imaging data, the next step is logically to access the temporal dimension. It is likely that the observed cell-to-cell variability is related to the fact that TAD formation is dynamic, especially if we think of TADs as ensembles of loops (see next section 1.2.1). In this sense the coming years will be extremely exciting as the techniques and the scientists engaged in deciphering TAD dynamics are improving and growing in number.

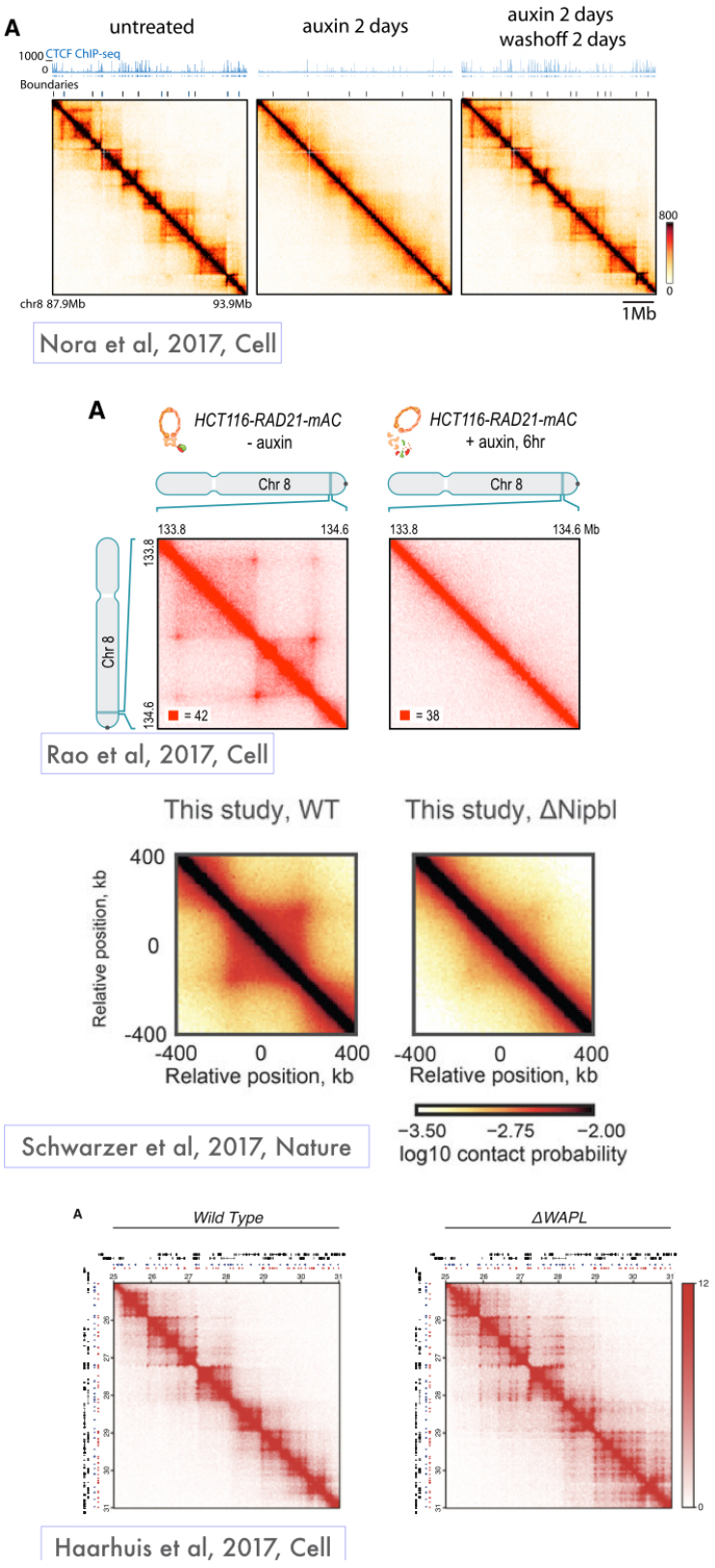



***Figure 1.13: TAD structure is lost if CTCF, Cohesin and Nipbl are degraded (first three from the top); long range interactions are enhanced when depleting Wapl (bottom).***

### 1.2.1 How are TAD formed: the loop extrusion hypothesis

The paradigm of DNA folding has significantly changed in the last 10 years. The advancements concern not only the observation of chromatin structure but also the modelling. The problem of describing and predicting chromatin arrangement in space has been tackled with different approaches. Researchers developed models based on pure polymer physics (Marko and Siggia, 1997) (Rosa and Everaers, 2008), others are based on epigenetic information to define preferential interaction (Jost et al., 2014) or on structural information (Giorgetti et al., 2014), or others built on DNA super-coiling (Benedetti et al., 2014).

For what concerns TADs, a proposed mechanism is gaining more and more attention and evidences in its support: the loop extrusion hypothesis. The basic idea is that specific proteins are capable of embracing the DNA string and form loops; as loops grow larger, distal regulatory elements are drawn near and a domain is formed.

Loop extrusion was first proposed in 1990 in a publication on the roles of DNA methylation (Riggs, 1990). In this work, Riggs named the mechanism "*DNA reeling*" and he foresaw the presence of, as he calls them, *Folding Elements*, proteins that bind DNA at specific sites, that move along chromatin and form loops, like restriction enzymes with no cleavage capacities [4]. He predicted that such a process would be ATP dependant, that folding was a solely *cis* process that structures could not arise from random interactions. ( Interestingly, the description of chromosomal loops that form on a scaffold resembles to a recent work on Mitotic condensation (Gibcus et al., 2018)). The idea of functional domains was already there (he talks about a 50-100 kbp scale) and in spite of the established idea of hierarchical folding (from histones, to the famous 30nm fiber, up to the solenoid) that we know as incorrect, I find it impressive how much Riggs could grasp of the mechanism that today seem to be the best candidate to describe TAD formation.

The idea of DNA loop as a basic mechanism of genome folding has been brought back in the early 2000s by Kim Nasmyth. The genetist dedicated a big part of his career to Cohesin and Condensin, two complexes respectively responsible for sister chromatid cohesion during DNA replication and chromosome condensation before cell division. In his dense review (Nasmyth,

[4] the reference to restriction enzymes is due to a previous work that proposed a sort of loop extrusion mechanism for restriction enzymes (Studier and Bandyopadhyay, 1988).

2001), Nasmyth suggests that Condensin could form DNA loops on top of a scaffold to resolve sister chromatids. He adds *"It is conceivable that Cohesin has a similar function"* (page 707 of the cited review).

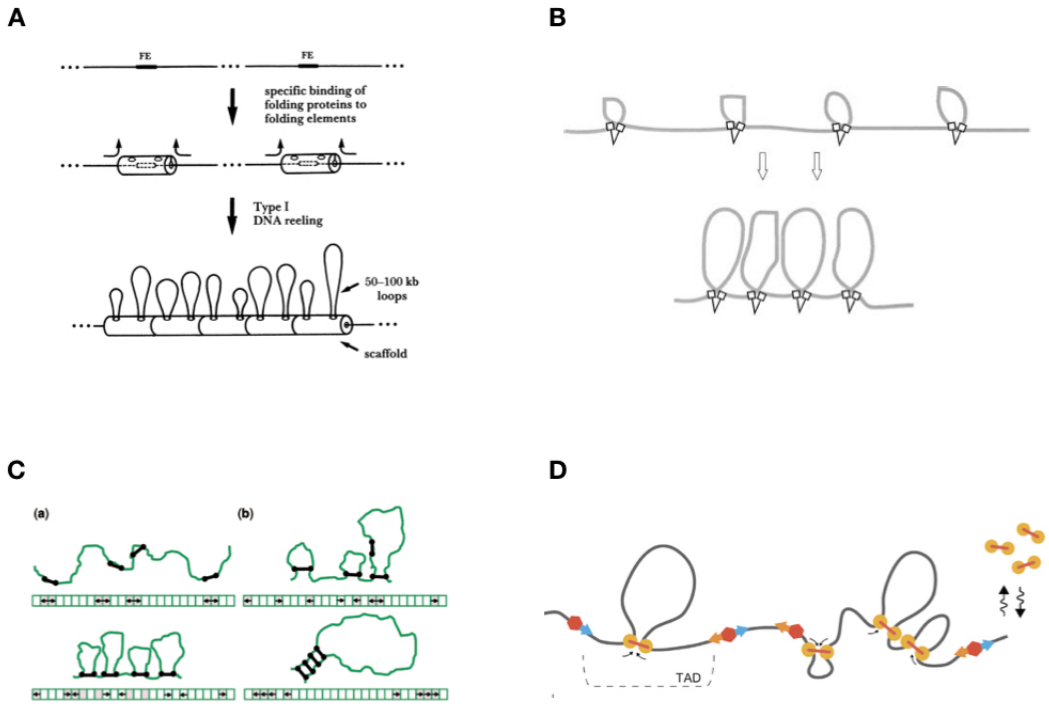



***Figure 1.14: The history of the loop extrusion hypothesis.*** ***A*** *The first proposed scheme for DNA reeling (i.e. loop extrusion) from a work on DNA methylation as a memory mechanism and genomic looping as a way to trigger functions, adapted from (Riggs, 1990).* ***B*** *Loop formation as a mechanism to separate sister chromatids, adapted from (Nasmyth, 2001).* ***C*** *A focus of loop extrusion energetics, again during Mitosis, adapted from (Alipour and Marko, 2012).* ***D*** *The loop extrusion hypothesis in its most complete and recent form. In this work a fully predictive model for interphase chromosomes is presented (Fudenberg et al., 2016).*

In 2012 Alipour and Marko published a detailed paper on the energetics of loop formation, that they name for the first time loop extrusion. They identify Cohesin as a candidate for *"(...) defining interphase chromatin loops."*,

hypothesis supported by two works that showed, a few years before, the role of Cohesin in insulation (Kagey et al., 2010) (Wendt et al., 2008).

A few years later a complete theory for loop extrusion is formulated (Fudenberg et al., 2016) (Sanborn et al., 2015). This model is capable of reproducing experimental data (i.e. Hi-C maps) for interphase chromosomal organisation. The ingredients are Loop Extruding Factors (LEFs) and Boundary Elements (BE) that stops the extrusion process and are positioned at TAD boundaries. The crucial parameters are the distance between LEFs (if two LEFs bump into each other they fall from DNA) and their *processivity*, or the amount of basepairs extruded while the LEF is bound. Once obtained the ensemble of conformations for chromatin they compute the corresponding simulated Hi-C map, which is comparable to the experimental one. Importantly, the loop extrusion model by Fudenberg and colleagues could predict the effects of biological perturbations (Schwarzer et al., 2017) (Nora et al., 2017).

The topic has gained more and more attention in the last 5 years and despite diverse proofs support the loop extrusion hypothesis there is still no direct evidence of such process *in vivo*. There is an *in vitro* demonstration of the extruding capacities of Condensin (Ganji et al., 2018).
The stakes in determining the mechanism behind chromatin dynamic folding is very high not only for its fundamental interest but also for the implications that architectural errors have. It has been largely shown that mis-arrangement of TADs is highly correlated to cancerogenesis (Valton and Dekker, 2016). In this context, the establishment of the paternity of the loop extrusion hypothesis has raised some debate. In this article the chronology of the basic concepts is briefly resumed (Dolgin).

## 1.3 CTCF and Cohesin: a focus on chromatin organisers

In this section I will introduce the objects of this study: CTCF and Cohesin. Fig 1.11 shows the main steps that have brought to the investigation of the interplay between CTCF and Cohesin; both factors where known to individually play some ”structural” role before the early 2000s, but no crosstalk was contemplated. In the last decade different groups demonstrated that CTCF and Cohesin are involved in the same processes (Kagey et al., 2010). In this section, after a brief description of each protein, I will present the basic

arguments that suggest their interaction in chromatin structure regulation.

### 1.3.1 CTCF

CTCF, the CCCTC-binding factor, is an 11-zinc finger, a DNA-binding protein with tens of thousands specific target sites on the genome (Ohlsson et al., 2001). CTCF was classically described as an insulator, a factor that bridges distal regulatory elements by looping DNA (Phillips and Corces, 2009) [5]. CTCF is also known to bridge interactions between specific genomic domains by marking Lamina Associated Domains boundaries (LADs)(Guelen et al., 2008), even though CTCF depletion does not affect these domains (Nora et al, unpublished results). Between 15,000 and 40,000 binding sites have been identified for CTCF, both at the boundaries and within TADs, and CTCF target sequence was found to be highly conserved during evolution (Kim et al., 2007). It has been recently shown that, in mammals, TAD boundaries are defined by CTCF convergent sites (Rao et al., 2014) (de Wit et al., 2015) (see also fig 1.19). Deleting or inverting these sites can have drastic consequences: already in one of the establishing works on TADs it is demonstrated that alteration of CTCF sites implies TAD disruption (Nora et al., 2012). Furthermore, Lupiañez and colleagues demonstrated that the alteration of some CTCF binding sites results in TAD mis-folding and consequent pathogenesis (Lupiáñez et al., 2015). Reshaping domains like TADs often means affecting gene expression, as perturbations of the insulation mediated by CTCF gives rise to enhancer-promoter interactions different from the wild type conditions. The correlation between TAD disruption and disease has been observed in different contexts (Valton and Dekker, 2016), CTCF sequence modification/deletion is generally a sufficient condition to affect chromatin compaction at the TAD scale.

In the following sections, and the rest of the manuscript, I will focus on mammals, and on mouse in particular.

### 1.3.2 Cohesin

Cohesin is a multiprotein complex. In its minimal description the Cohesin ring consists of two proteins belonging to the Structure Maintenance Complex family (Smc), Smc1a and Smc3, and a kleisin subunit, Scc1 also known as Rad21. Smc1a and Smc3 form a V-shaped heterodimer fused at what is known as the Hinge (see Fig 1.16); the two free arms, with their ATPase

[5] fig 4 of the reference (Phillips and Corces, 2009) nicely resumes the evidences and the hypothesis on the different roles of CTCF

| Name[a] | Gene | Species | Relationship to cohesin | Phenotype or disease | Reference |
|---|---|---|---|---|---|
| Scc4 | *MAU-2* | *C. elegans* | Loading factor | Axon guidance defect | Benard et al. (2004); Takagi et al. (1997) |
| Wapl | *WAPL* | *D. melanogaster* | Cohesin removal | Heterochromatin defect | Verni et al. (2000) |
| Scc1 | *RAD21* | *D. melanogaster* | Cohesin subunit | Axon pruning defect | Pauli et al. (2008) |
| Smc1 | *SMC1* | *D. melanogaster* | Cohesin subunit | Axon pruning defect | Schuldiner et al. (2008) |
| Scc1 | *RAD21* | *D. rerio* | Cohesin subunit | Haematopoiesis defect | Horsfield et al. (2007) |
| Smc3 | *SMC3* | *D. rerio* | Cohesin subunit | Haematopoiesis defect | Horsfield et al. (2007) |
| Pds5B | *PDS5B* | *M. musculus* | Cohesin regulator | Defects similar to CdLS | Zhang et al. (2007) |
| Scc2 | *NIBL* | *H. sapiens* | Loading factor | Cornelia de Lange syndrome (CdLS) | Krantz et al. (2004) |
| Smc3 | *SMC3* | *H. sapiens* | Cohesin subunit | Cornelia de Lange syndrome (CdLS) | Deardorff et al. (2007) |
| Smc1 | *SMC1L* | *H. sapiens* | Cohesin subunit | Cornelia de Lange syndrome (CdLS) | Deardorff et al. (2007) |
| Esco2 | *ESCO2* | *H. sapiens* | Cohesion establishment | Roberts SC phocomelia syndrome | Gordillo et al. (2008); Vega et al. (2005) |

[a] Generic human protein name.

***Figure 1.15: Table of Cohesin subunits and cofactors and diseases related to their alteration.*** *Adapted from (Wendt and Peters, 2009).*

domains, are bridged by Scc1. Cohesin's subunits and cofactors are listed in table 1.15.

The Cohesin complex is a big object measuring more than $\sim 50nm$ in height and up to $\sim 40 - 50nm$ of width (Anderson et al., 2002), as shown via Electron Microscopy in fig 1.16 right panel. There is no complete and final crystal structure for Cohesin, yet the left panel of fig 1.16 represents a likely scenario and gives an idea of the size of Cohesin compared to a nucleosome (between the two Smc arms). In this picture the reader can also appreciate the presence of other factors than those listed above: Pds5, Scc3, Wapl, Scc2/Scc4 are proteins needed for the proper functioning of Cohesin. Pds5 and Scc3 (SA1 and SA2 in mammals) are two additional components of the Cohesin complex, the so-called HEAT[6] repeated proteins Associated With Kleisins (HAWKs), or HEAT proteins associated with kleisins (Wells et al., 2017) (SA has been shown to be stably a part of the Cohesin complex (Gerlich et al., 2006)). From now on, when mentioning Cohesin I will refer to the heterodimer Smc1/Smc3 with the kleisin Rad21 and SA1/SA2.

Another protein interacting with the complex' keisins, but in a transient manner, is Scc2 (Nipbl). It has been shown that Nipbl is necessary for Cohesin loading on chromatin (Petela et al., 2018) it has also been suggested

[6] HEAT is an acronym that comes from 4 proteins: Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast kinase TOR1, (Andrade and Bork, 1995).

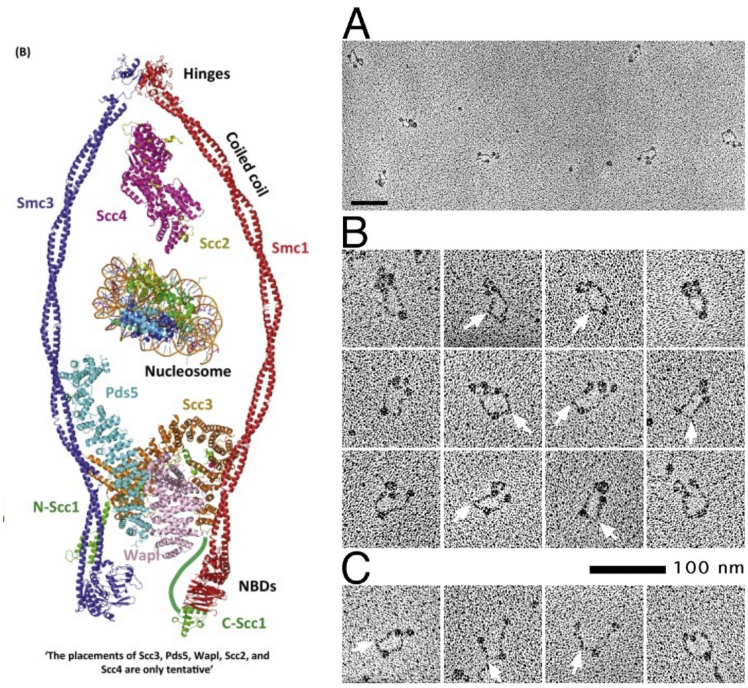



***Figure 1.16: Cohesin structure.*** ***Left*** *Temptative crystal structure for the Cohesin complex and related factors. Adapted from (Gligoris and Löwe, 2016).* ***Right*** *Human Cohesin complex captured with Electron Microscopy. Adapted from (Anderson et al., 2002).*

that once Cohesin loaded, Nipbl hops from one Cohesin complex to another to, most likely, stimulate its ATPase activity (Rhodes et al.). Wapl is commonly known as Cohesin's unloader, one of the major evidences for such claim being that Wapl depletion results in a *Vermicelli* (highly compacted chromatin) phenotype for interphase chromosomes (Tedeschi et al., 2013) (Wutz et al., 2017)[7]. The scenario in which Nipbl stimulates Cohesin ATPase activity and is in competition with Psd5 to bind the Scc1 subdomain, is plausible but not definite yet. A particular work showed that Pds5 and Wapl may promote loading of Cohesin on DNA in the absence of Nipbl (Murayama and Uhlmann, 2015) (a scenario that is contradictory, to say the least, given the many evidences of the effects of Wapl depletion (Tedeschi et al., 2013) (Wutz et al., 2017)). Nevertheless this discussed result is consistent with the

[7]It has actually been shown that the complex Scc2/Scc4 is required for the *Vermicelli* phenotype, as upon Scc4 and Wapl depletion chromatin was not condensed as reported in mutants lacking Wapl (Haarhuis et al., 2017)

fact that Pds5 depletion leads to TAD disruption (Wutz et al., 2017).

Until a few years ago Cohesin was mainly studied for its role in sister chromatid cohesion. Cohesin is the ring that keeps the two sister chromatids together during replication. In G1 Cohesin can perform transient binding, while in S phase a subpopulation of Cohesins is acetylated by the acetyltransferase Esco1 or Esco2. The acetylation locks Cohesin on chromatin, most likely to ensure proper cohesion during chromosome replication; the cohesion is finalised once Sororin has bound the complex. Sororin is a protein that competes for the Wapl binding site on Cohesin, its association to the complex is thought to have a stabilising effect. In Prophase, Sororin unbinds and most of the Cohesin is released from chromatin by the phosphorylation of the SA subunit. Meanwhile, Shugoshin (SGO1) and PP2A accumulates at centromeres to locally prevent Cohesins dissociation. Finally, in anaphase the leftover Cohesin rings are opened by a Separase that cleaves Rad21. The process is summarised in fig 1.17. For more details here are some publications on the topic: (Peters et al., 2008) (Nasmyth and Haering, 2009) Losada (2014) (Kanke et al., 2016).

There is still no consensus on the specific role of each factor involved in Cohesin metabolism. Recent evidences suggest that the picture may not be as binary as we think, proteins may not necessarily be accomplishing one simple function (loading, unloading, translocating, blocking) (unpublished results Nora). Two very recent works gave some insights on Cohesin structural remodelling upon association with DNA (Chapard et al., 2019) (Marko et al., 2018). What emerges is that Cohesin has two compartments, the one between the Smc arms and another one situated between the kleisin subunit (Rad21) and the ATPase heads of the Smc proteins. DNA is entrapped in the latter compartment as shown in both works summarised in fig 1.18. While in (Chapard et al., 2019) the problem is studied with an experimental, biochemical approach (Left panel in fig 1.18), the second reference (Marko et al., 2018) is the outcome of a theoretical model (right panel of fig 1.18).

Cohesin has also been partly characterised via *in vitro* single molecule assays (Stigler et al., 2016) (Davidson et al., 2016) (Kanke et al., 2016). The experiments presented in the three references are based on the same technique consisting in tethering many DNA strands on a coverslip, like a curtain, putting the protein of interest in solution and eventually apply a flow. DNA strands and the protein are detected in fluorescence, labelled with fluorophores of different colours; by doing so the behaviour of the protein, mainly residence time and sliding, can be characterised at the single molecule scale. It has been shown that Cohesin alone performs plenty of tran-

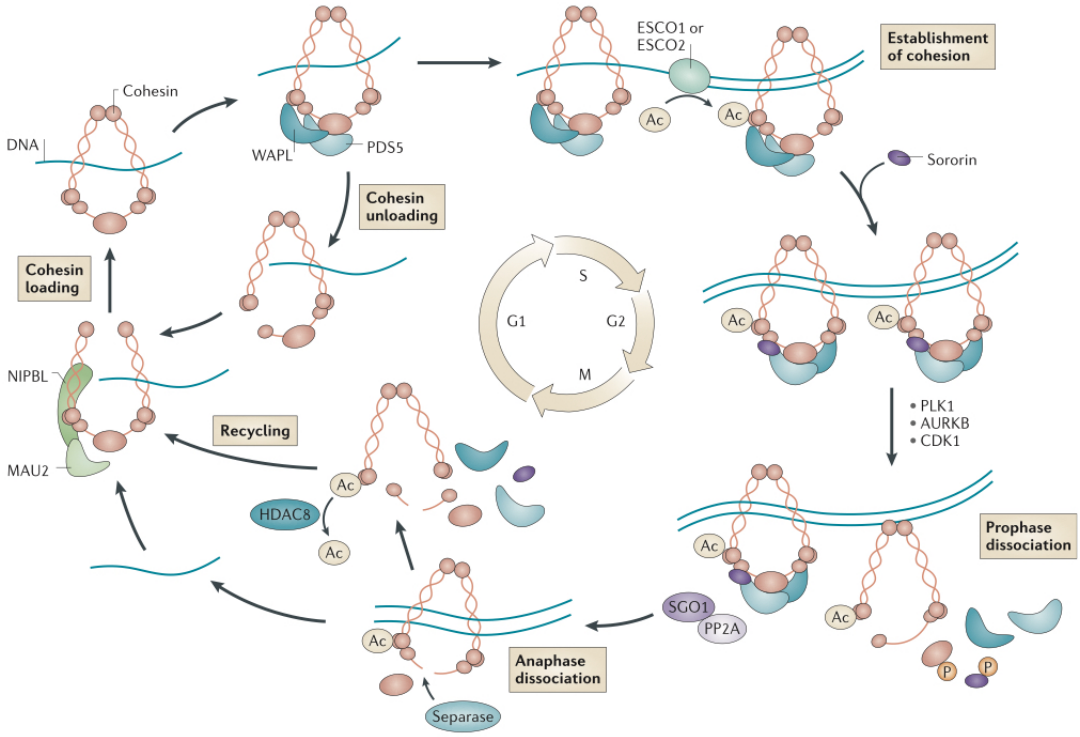



***Figure 1.17: Cohesin and the cell cycle.*** *Adapted from Losada (2014).*

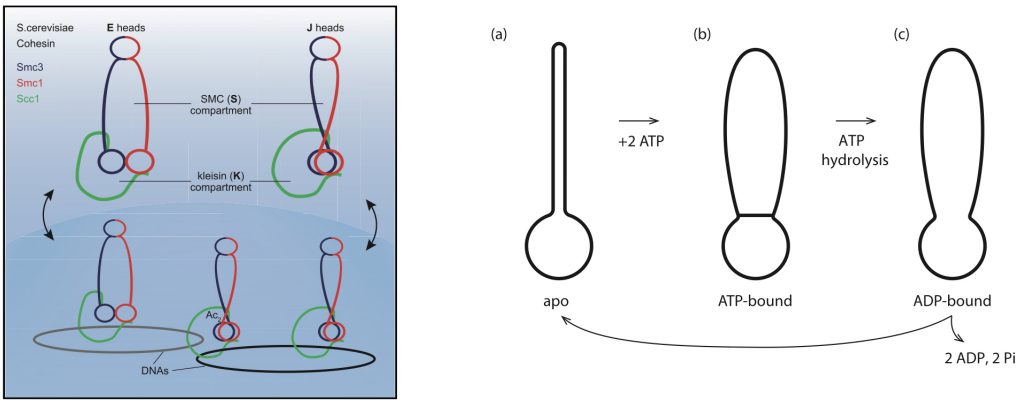



***Figure 1.18: Cohesin binding hypothesis Top*** *Adapted from (Chapard et al., 2019).* ***Bottom*** *Adapted from (Marko et al., 2018).*

sient interactions, that Cohesin sliding on DNA is ATP dependant. Stigler and colleagues reported that the presence of Cohesin loading factor Mis4 (the *S.pombe* homolog of human MAU2) the number of binding events were significantly increased, while Davidson et al showed that CTCF constrains Cohesin translocations. Both works showed that Cohesin is able of passing through many different obstacles (dCas9, EcoRI, Nucleosomes, and others). This beautiful results were obtained in an *in vitro* assay, mostly on naked DNA and in different salt conditions thus, despite their incredible value, they may not necessarily reflect what happens in a living cell. Most importantly, we do not know if Cohesin uses loop extrusion to move in these assays, or simply passively diffuses following the buffer flow exerted in the experimental setup.

For the sake of completeness I must mention that Cohesin is also involved in DNA repair during S and G2 phases (Wendt and Peters, 2009) but I will not treat the topic since this is the subject of a dedicated and complex field.

Based on the information described in this section I will focus in the following paragraph and further chapters on the role of Cohesin in the regulation of chromatin spatial organisation.

### 1.3.3 CTCF, Cohesin and chromatin structure

As mentioned at the end of the previous section, Cohesin is the ring that keeps the two sister chromatids together after DNA replication and for decades it

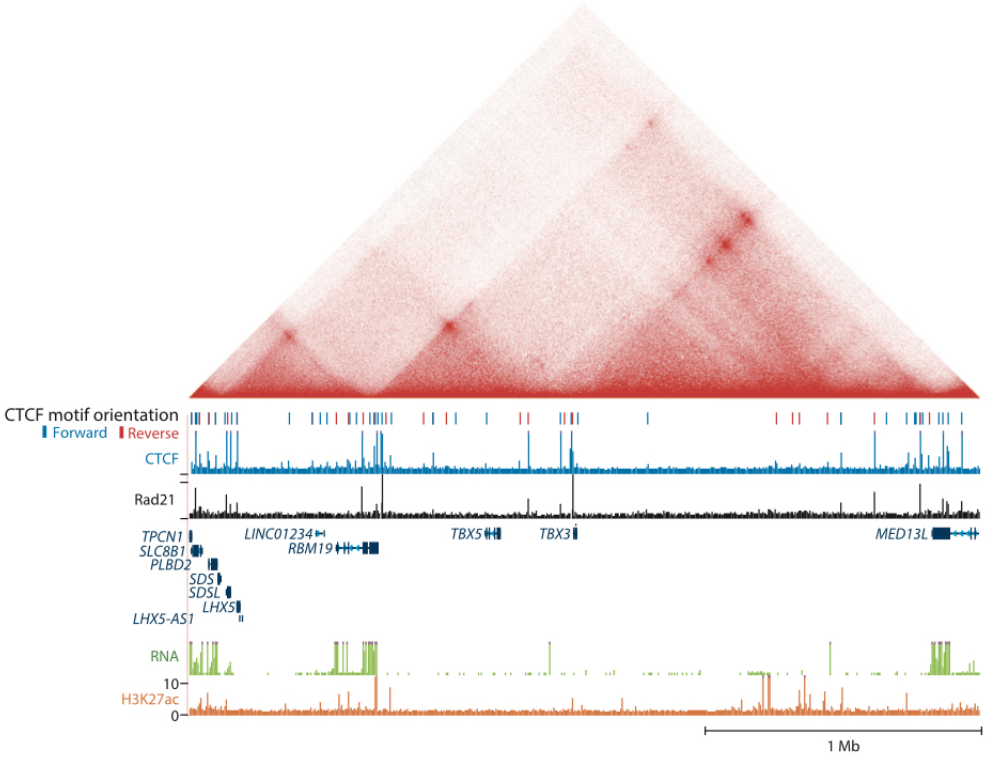



***Figure 1.19: CTCF and Cohesin are enriched at TAD boundaries.*** *ChIP-Seq profile of CTCF and Rad21, a subunit of the Cohesin complex, aligned with the Hi-C data from the GM12878 human lymphoblastoid cell line at the TBX5 locus. TADs can be recognised as the small triangles whose boundaries coincide with the ChIP-Seq peaks of CTCF and Cohesin. Adapted from (Merkenschlager and Nora, 2016)*

has been investigated in relation to DNA replication and cell division (Nasmyth and Haering, 2009). Still, its expression levels are considerable even in non-cycling cells, that is one of the reasons that prompted some scientists to explore other possible functions of the complex (Wendt et al., 2008). Soon, Cohesin has been found to play a role in gene regulation cooperatively with CTCF and with the Mediator complex in mammals (Kagey et al., 2010) and in *Drosophila* (Pauli et al., 2010).

Cohesin has been appointed as a possible DNA extruder, or LEF (Loop Extruding Factor) (or more in general involved in chromatin looping, see section 1.2.1), and various recent works have shown that it is fundamental to

maintain genome organisation at the TAD scale: (Wutz et al., 2017)(Gassler et al., 2017)(Rao et al., 2017)(Haarhuis et al., 2017). Interestingly, the same phenotype of disrupted TADs is observed when depleting CTCF (Nora et al., 2017). The effects of CTCF, Cohesin or Cohesin's co-factor depletion on the Hi-C maps are reported in fig 1.13. In addition, long range interactions take preferentially place at CTCF convergent binding sites and Rao and colleagues showed that Cohesin binds preferentially convergent CTCF binding sites[8] (Rao et al., 2014).
Taken together, all these evidences proof strikingly that both CTCF and Cohesin are involved in the regulation of chromatin structure at the TAD level.

How an extruding Cohesin may work is still unknown. It is not even clear if extrusion is performed by an individual Cohesin or by two or more. The group of Cees Dekker showed that *in vitro* Condensin extrudes only on one side (Ganji et al., 2018). It would be reasonable to think that Cohesin would behave similarly.

Despite the lack of direct evidence of loop extrusion, it is clear that chromatin organisation could not be a result of purely stochastic entrapment by Cohesin and it is even less likely that the driving mechanism is pure thermal motion. The need for convergent CTCF sites as a STOP sign for Cohesin, is a strong requirement. Indeed, alterations of the CTCF motif, inversion or deletion, lead to a loss of the related TAD (Lupiáñez et al., 2015).

## 1.4 Goal of this work

The aim of this work is to describe Cohesin dynamics and determine the nature of the interplay with other factors (i.e. CTCF, Sororin, Nipbl). The major findings in the field of chromatin architecture and its regulators are rely on static techniques, based on cell fixation, and the conclusions are issued from population averages. Our intention was to contribute with a single cell, single molecule approach. By choosing single molecule tracking we could study Cohesin in space and time, adding information about its dynamics.

[8] importantly CTCF target sequences are non palindromic.

# Chapter 2

# Technique and system

## 2.1 Optical microscopy

Optical microscopy is one of the most powerful tools developed in physics to study small living objects. Since its first conception, optical microscopy has been continuously evolving, driven by an ever growing curiosity to explore biological systems and processes in depth. After a brief historical introduction, I will give the fundamental elements of single molecule microscopy that, a part from providing a technical framework, motivate the choices made for the microscopy techniques in the context of this study.

Optical microscopy, or light microscopy, gets its name from the fact that it is used in the visible light wavelengths combined with a system of lenses to magnify a small object, or to quote Galileo "*un telescopio accomodato per veder gli oggetti vicinissimi*" ("An adapted telescope to look at very close objects", ed) (Saggiatore, 1623, Accademia dei Lincei, Roma). Fundamental contribution to the development of microscopes and to the study of living organisms came from van Leeuwenhoek and Hook, *Micrographia* (1665)), and a few decades before their wonderful achievements, Federico Cesi and Francesco Stelluti were taking advantage of the Galilean microscope to observe small organisms, such as honey bees. Their *Apiarium* (1625) is considered to be the first publication based on microscopy (see Fig. 2.1) [1]. This brief historical introduction stands as a proof of the tight relationship between optical microscopy and biology. Since the first *scopes* that helped revealing the details of entire organisms, huge technological improvements have been achieved and

[1] Sources: Grens K., 'Apiarium 1625', *The Scientist*, March $1^{st}$2015. University Library, University of Oklahoma, (galileo.ou.edu).

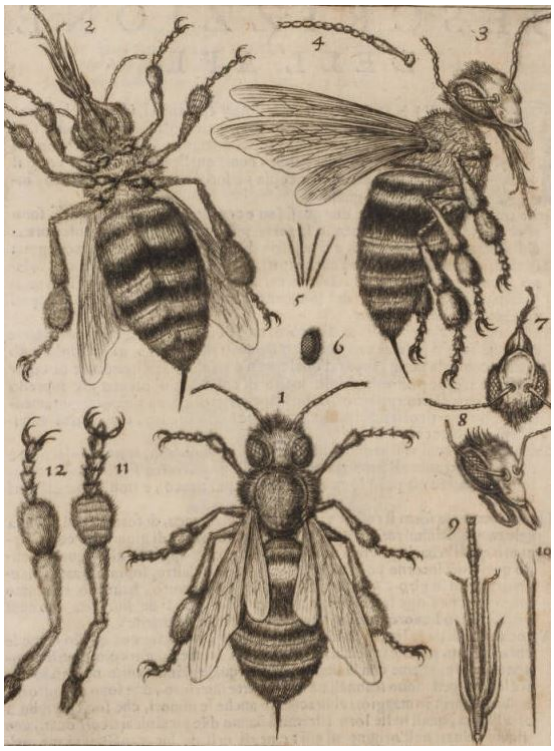



***Figure 2.1: Drawings of honey bees after observation with a microscope.*** *F. Cesi and F. Stelluti Apiarium,1625*

nowadays optical microscopy enables imaging down to the single cell and single molecule level.
In the following paragraphs I will detail the basics features of the microscopy technique chosen for this study.

### 2.1.1 Diffraction limit and fluorescence

Given the wave nature of light, optical microscopy is intrinsically limited by diffraction. A first quantification of the diffraction limited resolution was provided by Abbe in the nineteenth century; the diffraction limit can be estimated for the lateral and axial directions using the following equation:

$$d_{x,y} = \frac{\lambda}{2nsin\theta} \tag{2.1}$$

where $d$ is the radius of the imaged point, $\lambda$ is the wavelength of collection of the imaging lens, $n$ is the refractive index and $\theta$ is the half-angle of the light cone entering, or exiting, the lens. The Numerical Aperture of a lens

is defined as $NA = nsin\theta$. $NA$ is a dimensionless number that reflects the range of angles the light can be collected. One of the consequences of diffraction is that the image of an infinitesimal small point, the Point Spread Function (PSF), does not correspond to the physical size of the light source (see fig 2.2). As such, the diffraction limit basically determines the minimum distance needed to resolve two emitters placed in the close vicinity [2], as sketched in fig 2.2.

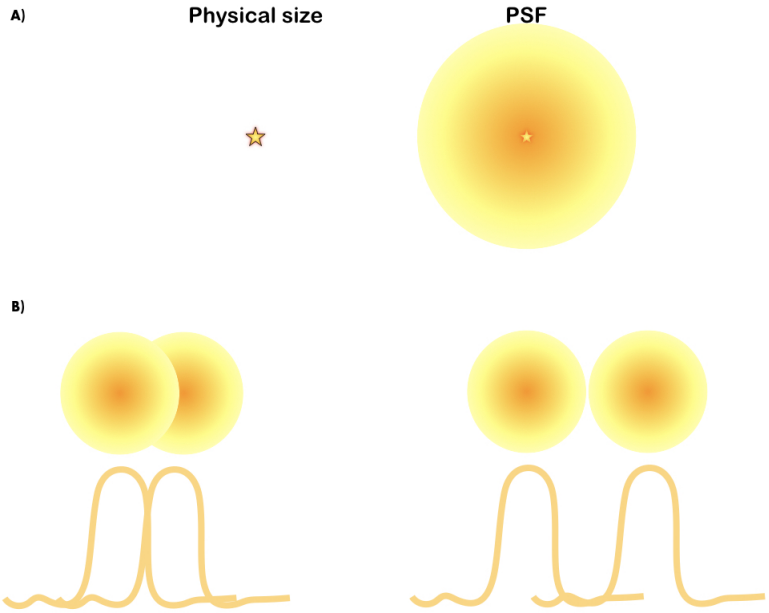



***Figure 2.2: Diffraction limited PSF. A*** *The actual size of the fluorescent probe visualised on the camera, for* $\lambda \sim 600nm$ *and* $NA = 1.4$ *is* $\sim 200nm$ *whereas the actual physical size of the probe is approximately 20 times smaller.* ***B*** *Given the diffraction limit, there is a minimum distance to resolve two point emitters.*

As stated in equation 2.1, the diffraction limit is determined by two factors: the wavelength and the numerical aperture. Great technological improvements have been done in the last years for what concerns objectives with high $NA$, but given the refraction index of biological samples the values is limited to $NA \sim 1.3$, in the context of single objective microscopy. A strategy to further improve the diffraction limited resolution could be to choose short wavelengths, but biological imaging in the last decades has re-

[2] This is usually referred as the Rayleigh criterion, for which the expression is the same as Abbe's but with a pre-factor of 0.61 instead of 2. I will not extend the discussion as the limiting quantities are the same: $\lambda$ and $NA$.

lied on optical microscopy combined to fluorescence in the visible spectrum. There are mainly two reasons why fluorescence is significantly favourable when imaging biological samples: first it allows specific labelling as fluorescent probes can be coupled to specific proteins, organelles and structures; second it is compatible with imaging in living cells.

The fluorescent probe chosen for this work is an organic dye, in particular the JF549 (Grimm et al., 2015). The dye was excited with a laser $\lambda_{ex} = 560nm$, and the emitted light was centred at $\lambda_{em} = 590nm$ due to Stokes shift. The objective has $NA = 1.45$ (see references in section 2.2) which is the maximal value for Total Internal Reflection Fluorescence (TIRF). As the imaging is performed in the nucleus of living cells, meaning at a depth $\geq 1\mu m$ in the sample, the angular capacity of the objective is smaller, and it can be estimated at $NA_{eff} \sim 1.4$. The resulting diffraction limit for the system presented in this manuscript is $d = \lambda_{em}/2NA_{eff} \sim 200nm$.[3]

Though diffraction limits the range of observable structures, imaging one single molecule at a time allows to localise its centre with a better precision compared to the PSF size. Other advantages are highlighted in the following paragraph.

### 2.1.2 Single Molecule tracking

To study proteins' dynamics I performed Single Molecule (SM)tracking experiments in living mouse Embryonic Stem Cells (mESCs). The main reason behind the choice of SM microscopy is that in contrast to other approaches often used to study particles' diffusive behaviour, such as Fluorescence Recovery After Photobleaching (FRAP) and Fluorescence Correlation Spectroscopy (FCS), SM imaging give access to the individual behaviour, avoiding ensemble averages. By doing so, different dynamic sub-population can be revealed; furthermore, looking at distributions, beyond mean values, helps maintaining the complexity and variability of many biological processes.
As stated in the previous section, there are particular requests to fulfil in order to localise and distinguish individual molecules.

To overcome the diffraction limited resolution, SM experiments rely on very low concentrations labelling, as shown in fig 2.3. Staining the cells with a nM or pM concentration of the fluorescent dye allows to assert that when a bright spot is detected, this spot is emanating from a single molecule. This "bright spot" correspond to the PSF defined in the previous section. If all aberrations have been corrected, the PSF is a 2D Airy function, commonly

[3] here $\lambda_{em}$ is the emission wavelength, as the imaging is performed in wide field fluorescence microscopy.

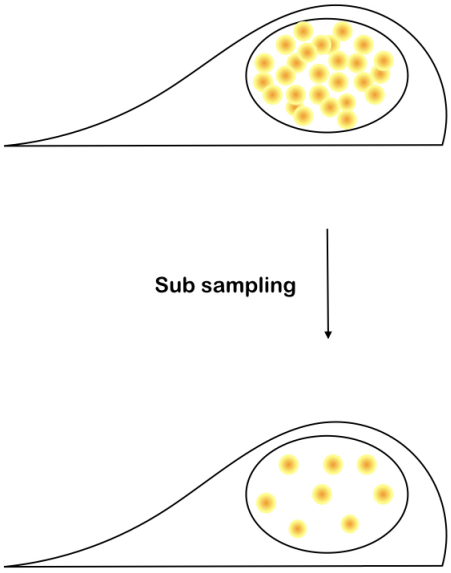



***Figure 2.3: Single Molecule live imaging is achieved via sub-labelling*** *To be able to follow one individual molecule we sub label the proteins bulk. To perform tracking we sampled* $\sim 10$ *molecules par frame.*

approximated to a 2D gaussian profile in the lateral dimensions (x, y) (see fig 2.2). By fitting the PSF with a 2D gaussian, it is possible to identify the centre of the PSF and localise the probe, with a precision better than tens of nm (assuming few thousands of emitted photons per molecule) which is less than ten nanometers in size (such as an organic dye), see fig 2.4.

A crucial factor to perform a good single molecule imaging experiment is the Signal to Noise ratio (S/N). As explained in (Thompson et al., 2002), the limiting factor for single molecule detection is the number of photons collected with respect to the background noise. In fact, the localisation precision is defined by $\sigma = s/\sqrt{N}$, where $s$ is the size of the image and $N$ is the number of photons collected. Increasing the number of photons is a strategy to improve the localisation precision. Nevertheless there are other factors to be considered for an optimal single molecule experiments. A crucial quantity is the Full Width at Half Maximum (FWHM) of the PSF both in the later (x,y) and the axial (z) directions, which can be computed with the following expressions:

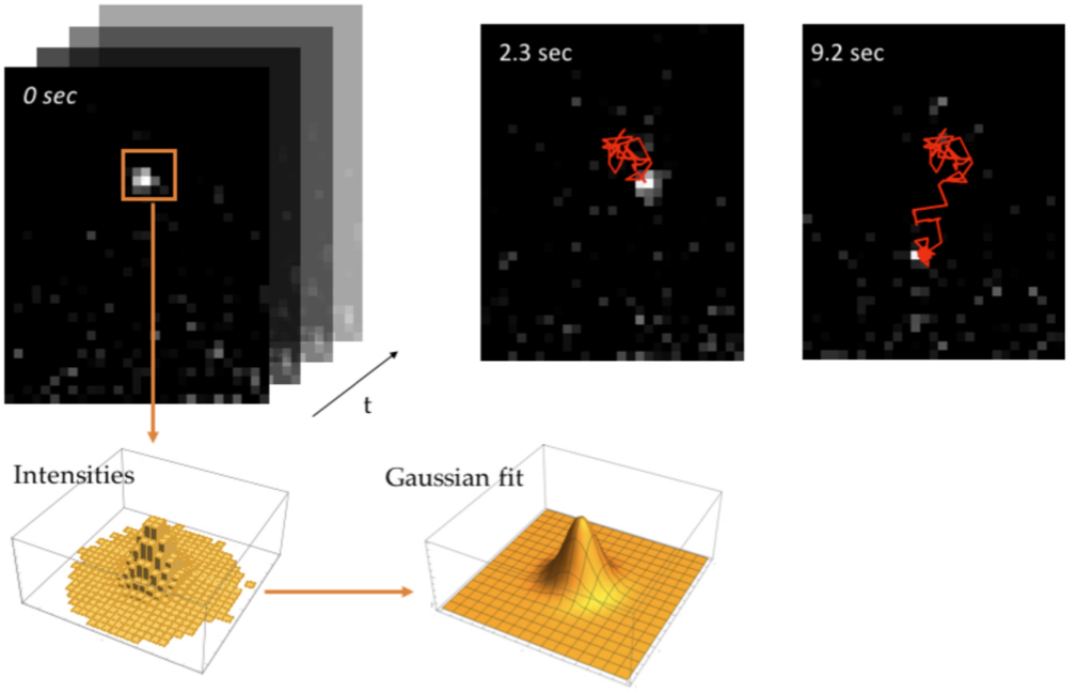



***Figure 2.4: Single Molecule localisation and tracking.***

$$FWHM_{xy} = \frac{\lambda}{2NA} \qquad (2.2)$$
$$FXHM_z = \frac{2\lambda}{NA^2}.$$

As shown in fig 2.5, the FWHM along the axial position corresponds to the Depth of Focus of the imaging set up. In the case of this study, given an emission wavelength $\lambda em \sim 600nm$ and an effective numerical aperture $NA_{eff} = 1.4$, the DOF is $\sim 600nm$. Only the molecules within a sheet of $600nm$ will appear in focus are properly localised and yet, as the entire specimen is illuminated, all the molecules will be contributing to the background. The aim of a single molecule microscopist is to optimise the signal to noise ratio thus, with this goal in mind, it is crucial to keep the background as low as possible. For this reason a considerable effort is put in the optimisation of experimental parameters such as laser power and exposure time. Importantly, the latter parameter is set not only by the photo-physics of the fluorophore

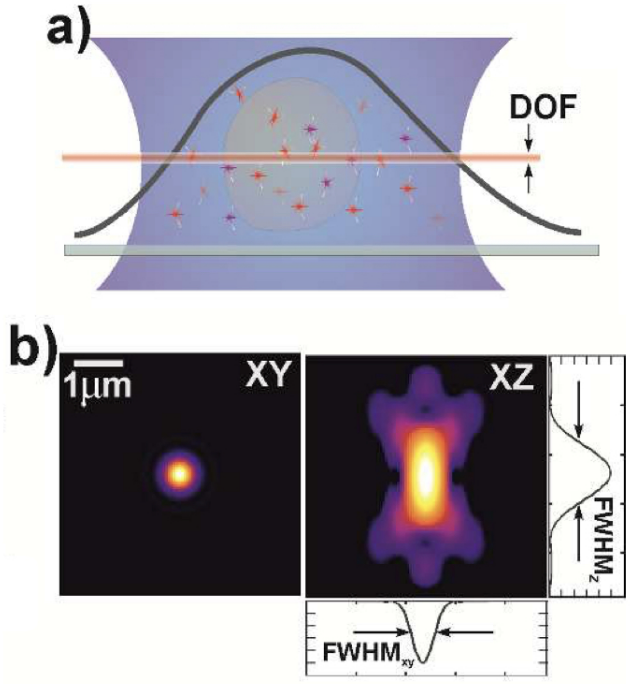



***Figure 2.5: Sketch of Single Molecule optical microscopy.*** ***a*** *Wide-field imaging: the whole specimen is illuminated but only the volume within the Depth Of Field (DOF) is imaged. All the molecule outside the focal plane will contribute as background.* ***b*** *Point Spread Function (PSF) of a single emitter in an aberration-free wide-field optical microscopy, both from the lateral (left) and axial (right) point of view. Adapted from (Hajj et al., 2014).*

but also by the characteristic time-scale of the process of interest. As briefly mentioned at the beginning of the section, the work presented here is based on imaging of mESCs. Stem cells are very thick and they tend to grow in colonies, often on top of each other. In addition, stem cells are extremely sensitive to light and temperature. For these reasons the optimisation of the crucial experimental parameters mentioned above, was quite challenging.
Shining light in the conventional wide field configuration gave very poor results in terms of Signal to Noise ratio (S/N). To limit the considerable background coming from such volumetric cells we adopted a microscopy approach called Highly Inclined and Laminated Optical sheet, or HILO (Tokunaga et al., 2008). Combining the inclination of the beam, as shown in fig 2.6 panel a, and a a field stop placed in the plane conjugated to the specimen to limit the illuminated portion of the nucleus, fig 2.6 panel b, the S/N was

significantly improved (see Fig 2.6, bottom). Beyond the considerable gain in terms of S/N, choosing HILO improved the cell viability as the amount of excitation light required to obtain a sufficient S/N is limited.

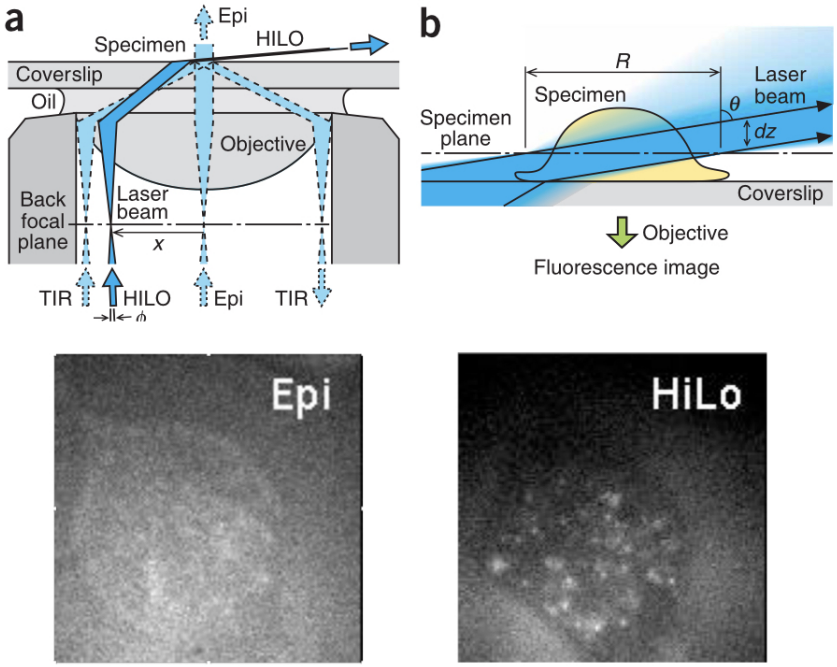



***Figure 2.6: HiLO microscopy and its advantages Top: (a)(b)*** *Scheme of HILO illumination, adapted from (Tokunaga et al., 2008).* ***Bottom*** *Snapshots of a single nucleus in Epifluorescence (or wide field) configuration (left) and in HILO configuration (right). Images taken by Antoine Coulon.*

## 2.2 Microscopy set-up

Single molecule imaging was performed on an epifluorescence inverted microscope (IX71, Olympus) in Highly Inclined and Laminated Optical sheet, or HILO illumination. Proper HILO was achieved using a slit (or field stop) placed in the plane conjugated to the specimen plane (OWIS, 14.021.0020, RT 40-20-R). Lasers beams were focused in the back focal plane of a 150X objective lens (UAPON 150XOTIRF, Olympus, France), selected by an excitation quadband dichroic supplied with the corresponding emission filters (Chroma, TRF89901-EMv2 - ET - 405/488/561/640nm Laser Quad Band

Set for TIRF application). Lasers were tuned via an acousto-optical tunable filter (AOTFnC-400-650-TN, A&A Optoelectronic, France) and controlled by a home-made interface in Micromanager (Edelstein et al., 2014). Laser power was adjusted to have a density of $\sim 0.1 kW/cm^2$. Signal was acquired with an EM-CCD camera (iXonEM DV860DCS-BV, Andor, Ireland) run in frame transfer mode. The setup is provided with a 405 nm laser (Cube 405- 100C, Coherent, Santa Clara, CA, USA), a 488 nm laser (35LAL030-220, CVI, Melles-Griot, France) and a 561 nm laser (Genesis MX 561-2000 MTM, Coherent, Santa Clara, CA, USA).

## 2.3 Biological system

To perform live single molecule imaging an Halo-tag was encoded in the protein sequence. Thanks to the tag I could label individual proteins and follow them around the nucleus.The cell lines were edited by Elphège P. Nora in the laboratory of Benoit Bruneau at the Gladstone Institute in San Francisco, California (USA).

Cohesin has been tracked in the context of various alterations. In particular I studied how Cohesin dyamics is affected if we deplete factors like CTCF, Sororin and Nipbl. Figure 2.7 reports a sketch of the edited cell line where I tracked Cohesin in the absence of CTCF and the concept of the auxin inducible degradation system is represented.

The different conditions, and the corresponding cell lines are listed in the table reported below.

| **what is edited** | **cell line ID** | **tissue** | **goal** |
|---|---|---|---|
| CTCF-Halo | EN129.2 | mESC (S/G2) | Tracking CTCF WT |
| Cohesin-Halo WT | EN130.1 | mESC (S/G2) | Tracking Cohesin WT |
| Cohesin-Halo CTCF-AID | EN131.1 | mESC (S/G2) | Tracking Cohesin +/- CTCF |
| Cohesin-Halo CTCF-AID | EN228.2 | Astrocytes (G0) | Tracking Cohesin +/- CTCF. |
| Cohesin-Halo Sororin-AID | EN229.3.1 | mESC (S/G2) | Tracking Cohesin +/- Sororin |
| Cohesin-Halo Nipbl-AID | EN273.3 | mESC (S/G2) | Tracking Cohesin +/- Nipbl |
| CTCF-2A-AID-Halo-NLS | EN132.1 | mESC (S/G2) | Control: Halotag-dye and NLS |

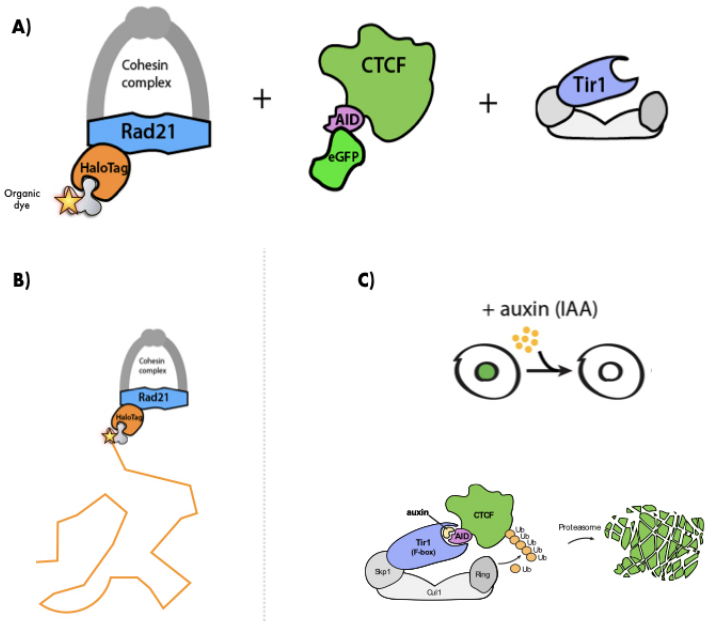



***Figure 2.7: Description of the cell lines.*** ***A*** *sketch of CRISPR edited mESC stable cell line expressing Cohesin-Halo in the background of CTCF-AID.* ***B*** *Each Cohesin can be individually labelled by coupling it with an organic dye fused to the Halo Ligand.* ***C*** *Sketch of CTCF depletion via the degron system upon incubation with auxin. Partially adapted from (Nora et al., 2017).*

.

All the insertions at the proteins sequences (CTCF and Rad21 for Cohesin) are are at the endogenous gene encoding for these proteins (not transgene over-expression), except for the Tir1 box necessary for the degron system which is not on both alleles and was randomly integrated in the CTCF-AID line, as published before (Nora et al., 2017).
The control cell line EN132.1 has been produced from the endogenous CTCF locus but separated from the CTCF protein by a 2A self-cleaving peptide. Gene editing in mESCs was performed as described in (Nora et al., 2017). CRISPR sgRNA sequences were cloned into the Cas9 nuclease encoding vector pX330 (Addgene #42230), except when targeting the CTCF N-terminus

where the Cas9 nickase (D10A) encoding vector pX335 (Addgene #42335) was used. Corresponding targeting vectors are listed below. FRT flanked selection cassettes were removed by transient transfection of a Flippase encoding vector and mESC subcloning, as described in Nora et al. 2017. Successful targeting was confirmed by genotyping genomic DNA by PCR and when appropriate Western blot as well as flow cytometry and microscopy to confirm expression levels and nuclear localization of fusion proteins. Annotated sequences of targeting vectors will be provided through Addgene upon publication, and are available upon request.

| **Cell line genotype** | | |
|---|---|---|
| WT parental | 19 XY | - |
| CTCF-AID-eGF Tir1(random integration) | homozygous except for Tir1 | clone1 |
| Rad21-Halotag | homozygous | clone1 |
| Rad21-Halotag, CTCF-AID-eGFP Tir1(random integration) | homozygous except for Tir1 | clone1 |
| Rad21-Halotag, CTCF-AID-eGFP Tir1(Tigre) | homozygous | clone1 |
| Rad21-Halotag, Sororin-AID-eGFP Tir1(Rosa26) | homozygous | clone1 |
| CTCF-AID-Halotag | homozygous | clone1 |
| CTCF-2A-Halotag | homozygous | clone1 |
| eGFP-AID-Nipbl, Tir1(Tigre) | homozygous | clone1 |

| **targeting vectors** |
|---|
| pCAGGs-Flpo-IRES-puro |
| pEN113 - pCAGGS-Tir1-V5-BpA-Frt-PGK-EM7-NeoR-bpA-Frt-Rosa26 |
| pEN114 - pCAGGS-Tir1-V5-BpA-Frt-PGK-EM7-PuroR-bpA-Frt-Rosa26 |
| pEN244 - CTCF-AID[71-114]-eGFP-FRT-Blast-FRT targeting construct |
| pEN313 - Rad21-Halo-Frt-PGK-EM7-NeoR-bpA-Frt targeting |
| pEN372 - CTCF-2A-3Xnls-AID[71-114]opt-Halo-Frt-PGK-EM7-NeoR-bpA-Frt targeting |
| pEN396 - pCAGGS-Tir1-V5-2A-PuroR TIGRE donor |
| pEN487 - Sororin-AID[71-114]-eGFP-FRT-Blast-FRT targeting construct |
| pEN84 - CTCF-AID[71-114]-eGFP-FRT-Puro-FRT targeting construct |

### 2.3.1 Cell Culture

Mouse Embryonic Stem Cells (mESC) were cultured in DMEM+Glutamax (ThermoFisher cat 10566-016) supplemented with 15% Fetal Bovine Serum (DUTSCHER Ref S1810-050 Lot S15642S1810), 550mM b-mercaptoethanol

(ThermoFisher 21985-023), 1mM Sodium Pyruvate (ThermoFisher 11360-070) and 104U of Leukemia inhibitory factor (Millipore ESG1107). Cells were maintained at a density of $0.2 - 1.5 * 10^5$ cells $/cm^2$ by passaging using TrypLE (12563011) every 24-48h on 0.1% gelatin-coated dishes (Millipore cat ES-006-B) at 37° and 5% CO2. Medium was changed daily when cells were not passaged. Cells were checked for mycoplasma infection every 3-4 months and tested negative.

Neural Progenitor cells were cultured in N2B27, medium composed by: DMEM/F12 (Gibco 31330-038) and Neurobasal medium (Gibco 21103-049) supplemented with 5mL L-Glutamine 100X (Gibco 25030-024), 10mL B27 (50X) (Gibco 17504-044), 5mL N2 (100X) (Gibco 175020-01),
2mL 2-mercaptoethanol (50mM) and 10ng/mL EGF and FGF (Peprotech). Cells were cultivated in gelatin coated dishes and passed every 2-3 days using Accutase. For differentiation into quiescent Astrocytes NPCs were washed from N2B27+EGF+FGF and cultured for 48h in N2B27 supplemented with 10 ng/mL BMP4 (R&D Systems).

## 2.4 The experiments

To perform single molecule tracking experiments, cells (both mESC and Astrocytes) were grown on circular petri dishes with glass bottom (MatTek, Part No: P35G-1.5-14-C) preventively coated with fibronectin (Millipore SAS cat FC010-5mg). Cells were seeded at a density of $3 * 10^5/cm^2$ the day before the experiments in Fluorobrite DMEM (Life Technologies SAS - Thermo Fisher Scientific cat A1896701).
I underline the importance of performing single molecule imaging in phenol-red free medium to reduce the background fluorescence.

### 2.4.1 The degron system

To investigate the interplay between Cohesin and other factors such as CTCF, Sororin and Nipbl, the first has been tracked in presence and absence of the latter. Protein depletion has been achieved via a novel and very efficient system called *degron* (Nishimura et al., 2009).

The degron system is based on a degradation pathway that is responsive to the hormone auxin in plants. Mammalian cells do not have the auxin-responsive machinery but share the degradation pathway with the plants; thus, by integrating an auxin inducible cassette, known as the Tir1 box and by adding the auxin responsive sequence to the protein of interest, it

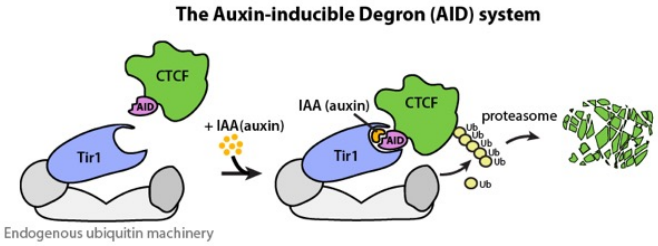



***Figure 2.8: The degron system.***

is possible to achieve degradation upon incubation with the hormone. In presence of auxin the auxin Inducible Degron (AID) couples with the Tir1 box, which is responsible for the recruitment of Ubiquitines that induce the protein degradation. A sketch of the degron system is reported in fig 2.8.

Classical strategies of proteins depletion rely on alteration of the coding sequence or interference with the mRNA via interfering RNAs (RNAi). In the first case the modification is non-inducible but permanent, and may affect dramatically cells viability, as proteins that are essential for development, mitotic checkpoint or many other fundamental processes cannot be permanently deleted from the genome. On the other hand, degradation of the mRNA does not necessarily imply a complete depletion or the targeted protein, and depends on its stability.

A striking example of the artefacts due to a non-fully efficient degradation is provided by two works based on CTCF depletion. In (Zuin et al., 2014) CTCF is degraded via RNAi and the effects on chromatin insulation are mild; recently, Nora showed that degron depletion TAD are completely lost (Nora et al., 2017). The main reason behind such discrepancy is due to the fact that RNAi degradation still leaves 10-15% of the targeted protein (Zuin et al., 2014) and, as shown by Nora by titration of the amount of remaining protein with respect to auxin concentration, strong effects are visible when just $< 4\%$ of CTCF is left (see fig 2.9).
Another important aspect of the degron system is its kinetics. As shown via fluocytometry, it takes 2 hours of incubation with auxin to achieve complete depletion of the target protein, see fig. 2.10. Furthermore the degron system is reversible: if auxin is washed off the expression levels go back to the wild type condition.
Last, but not least, the degron system ensures homogeneity in the cell population in terms of degradation. As shown by fluocytometry and confirmed by our experiments, after 2 hours the population reaches a stationary state and show uniformity in the depletion levels.

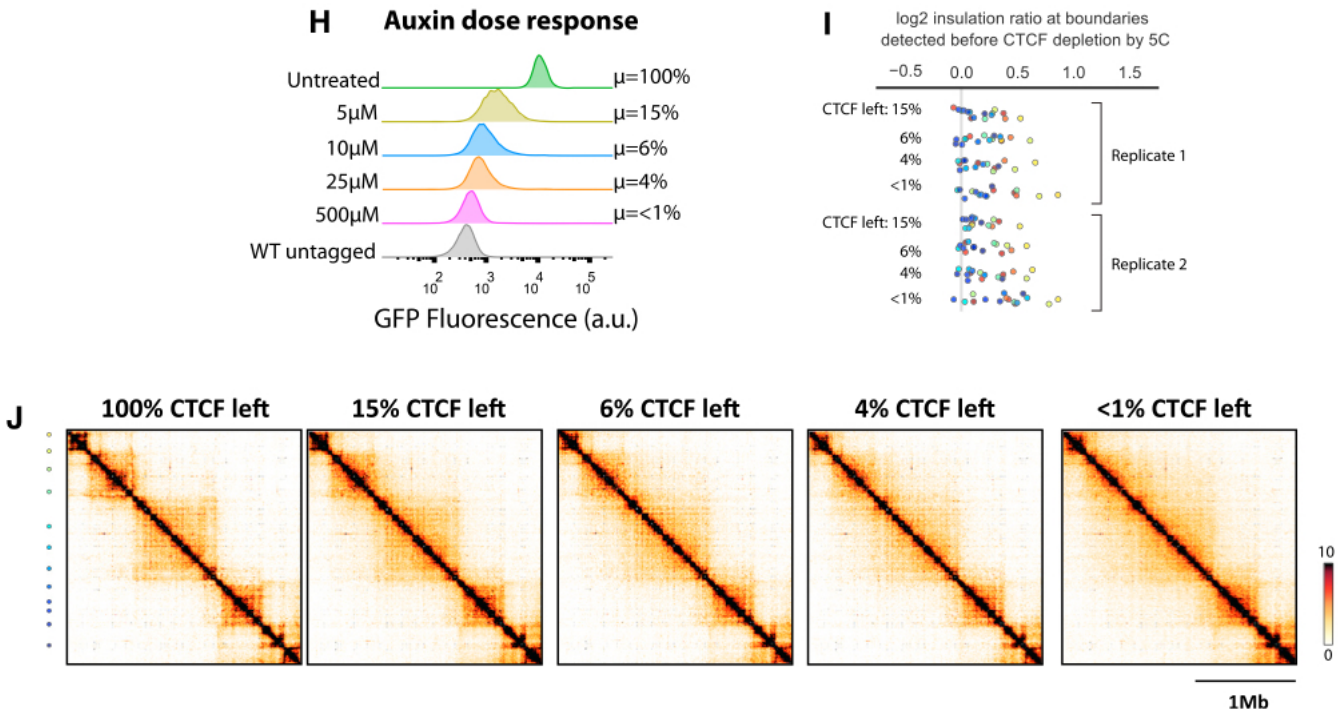



***Figure 2.9: A full depletion of CTCF is needed to observe a real effect on chromatin insulation.*** *Panel H: Titration of the CTCF leftovers as a function of auxin concentration. Panel I: quantification of the insulation level as a function of CTCF concentration. Panel J: 5C maps for different amount of CTCF left in solution; $>$ 4% of CTCF is enough to maintain some insulation. Adapted from (Nora et al., 2017).*

### 2.4.2 Labelling and imaging conditions

To trigger proteins degradation via the degron system I added auxin to cell culture medium (IAA-Indole-3-acetic acid sodium salt ref : I5148-2G, Sigma-Aldrich). In Fig.2.10 is reported the kinetics of depletion measured via fluo-

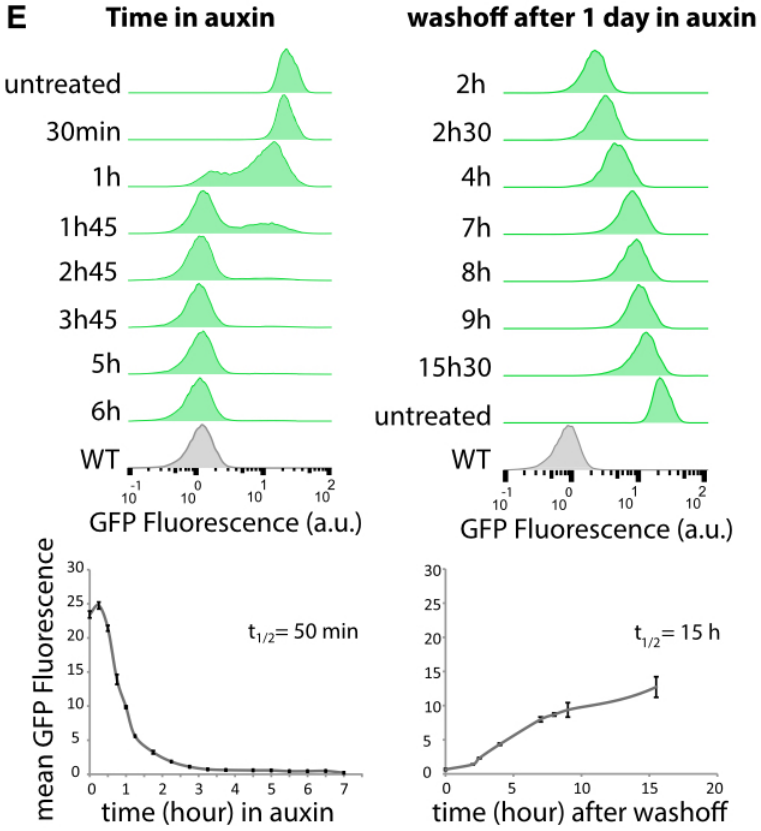



***Figure 2.10: Kinetics of CTCF depletion measured via fluorocytometry.*** *Adapted from (Nora et al., 2017).*

rocytometry.

To achieve single molecule labelling cells were incubated with 1pM of Halo-JF549 for 20 minutes at 37° (incubation followed by a first rinsing step, 15 minutes wait and another rinsing). While waiting for the second rinsing step cells were incubated with 1uM Hoechst and consequently washed to minimise the fluorophores unbound in solution. All washings were performed using cell culture medium; the coverslips treated with auxin were washed with medium enriched with auxin. During the experiments cells were kept at 37° and 5% CO2 with a Tokai Hit heating system (INUBG2E-PPZI).

To precisely identify the nucleus, cells were stained with $1\mu M$ Hoechst 33342 (bisBenzimide H 33342 trihydrochloride, Sigma-Aldrich, ref 14533),

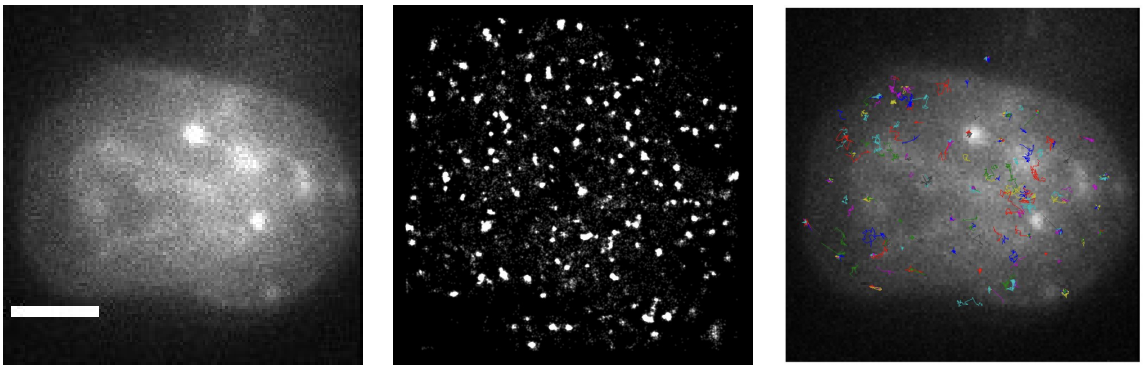


***Figure 2.11: Extracting trajectories from single molecule localisations.*** *From left to right: image of a nucleus in the Hoechst channel; maximum projection of all localisations obtained by tracking Cohesin-HaloJF549; trajectories obtained from the localisations (filtered for length ¿ 10 frames) merged with the Hoechst image. Scale bar = 5 µm.*

excited with 405 nm light. The presence/absence of CTCF-GFP was revealed in the 488 nm channel. A snapshot was systematically taken in both these two channels. To track Cohesin-Halo-JF549 the sample was excited with the 561 nm laser. Movies were recorded in a continuous imaging regime, the laser being controlled by the camera.

## 2.5 Analysis of single molecule imaging

One single dataset corresponds to the pool of trajectories obtained from a single cell, which is of the order of thousands trajectories. For each biological condition 10-15 cells were imaged.

To localise the single emitters and build the trajectories we used a home-made software (SLIMFast, (Normanno et al., 2015)) implemented in Matlab and based on the MTT algorithm (Sergé et al., 2008). The Point Spread Function (PSF) of a single emitter is fitted with a 2D-gaussian, whose center corresponds to the position of the fluorophore with a sub-pixel resolution.

To exhaustively characterise CTCF and Cohesin dynamics, I performed experiments at different acquisition rates. This approach enabled me to overcome the bleaching limit and cover different timescales. I performed acquisitions with an exposure of 5ms, 50ms and 500ms; when increasing the exposure time the laser power was reduced coherently (keeping the Signal to

Noise qualitatively constant).
Data from experiments at 5ms exposure were used to quantify the fraction of bound molecules, since at this rate there is no bias towards one of the dynamic subpopulations. In fact at slower rates highly mobile proteins are blurred and often not properly localised.
Data from experiments at 50ms served to characterise the dynamics via the computation of the Mean Square Displacement (MSD) and the binding kinetics, with the residence time distribution, or Survival Probability.
Data from experiments at 500ms were used to quantify the binding kinetics on longer time-scales. At such rate I was indeed completely biasing the acquisitions and the analysis towards stably bound molecules.
The methods used to analyse the trajectories will be explained in detail in the following paragraphs.

### 2.5.1 Analysis of binding kinetics

To quantify the number of bound molecules I exploited trajectories from the acquisition at 5ms. By doing so I could include all the trajectories, even the shortest ones consisting of only 2 displacements.

The analysis was performed with SpotOn (Hansen et al., 2018). The method is based on a fit of the distribution of all the step lengths performed by the molecules. The idea behind is that proteins can be bound or freely diffusing. The state of the protein (bound or unbound) is reflected in the distribution of the steps that it perform: a bound molecules will give raise to small steps while a freely diffusing one will show longer steps. The kinetic modelling is inspired by (Mazza et al., 2012). The so-called jump distribution is resumed by the following expression:

$$P(r,\Delta t) = F_{bound}\frac{r}{2(D_{bound}\Delta t+\sigma^2)}e^{-\frac{r^2}{4(D_{bound}\Delta t+\sigma^2)}} + (1-F_{bound})\frac{r}{2(D_{free}\Delta t+\sigma^2)}e^{-\frac{r^2}{4(D_{free}\Delta t+\sigma^2)}} \tag{2.3}$$

The software provides a correction for motion in the axial direction. Particles can indeed be easily lost since they explore the space in 3D but our imaging is performed on a 2D projection. Hansen and Woringer introduced a correction based on the computation of the probability that a molecule leaves the focal plane, based on its diffusion coefficient and as a function of time. Both the fitting model and the correction are based on the assumption of pure Brownian motion. Such a choice is arguable since the dynamics observed is often sub-diffusive, nevertheless since this kind of analysis is based

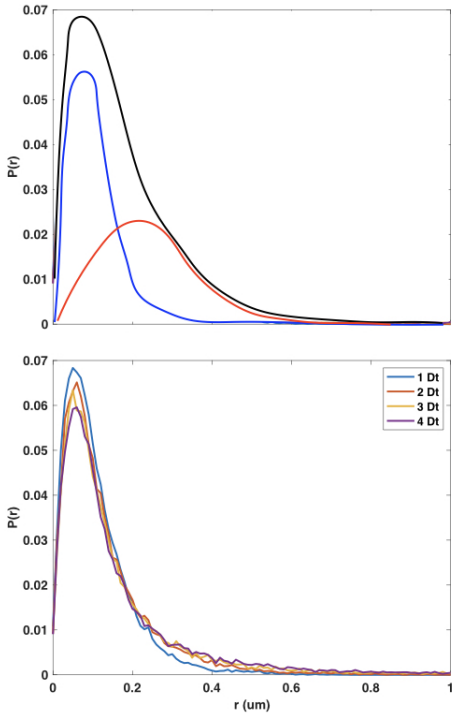



***Figure 2.12: Step length distribution and diffusive sub-population: how I infer the fraction of bound molecules.*** ***Left:*** *an example of the distribution of step lengths, P(r). The Blue curve correspond to the distribution arising from bound molecules that perform short steps. The red curve represents the distribution of freely diffusing molecules, with longer steps. The black curve mimic the global distribution, sum of the two sub-populations. textbfRight: An example of empirical distribution for Cohesin in wild type conditions. The 4 different curves represent the distributions for increasing time lags, from 1 to 4* $\Delta$ *t.*

on very short trajectories (mean step of 2-3 frames and median of 4-5) the diffusive assumption is a fair hypothesis.

A two-state model was chosen to fit our data, corresponding to the scenario illustrated in fig 2.12. For each data set I computed the fraction of bound molecules, or $F_{bound}$, via a fit that was performed on the Cumulative Distribution Function (CDF) to avoid biases due to the binning choice. To extract a characteristic value of $F_{bound}$ for a specific biological condition, I averaged the results obtained par dataset (= par cell) and extrapolated the corresponding standard deviation.

### 2.5.2 Residence time

To further characterise the binding kinetics, I quantified the proteins residence time distribution. To do so, I extrapolated the trajectories that stayed confined in a circular area for the whole duration and considered them as stably bound molecules.

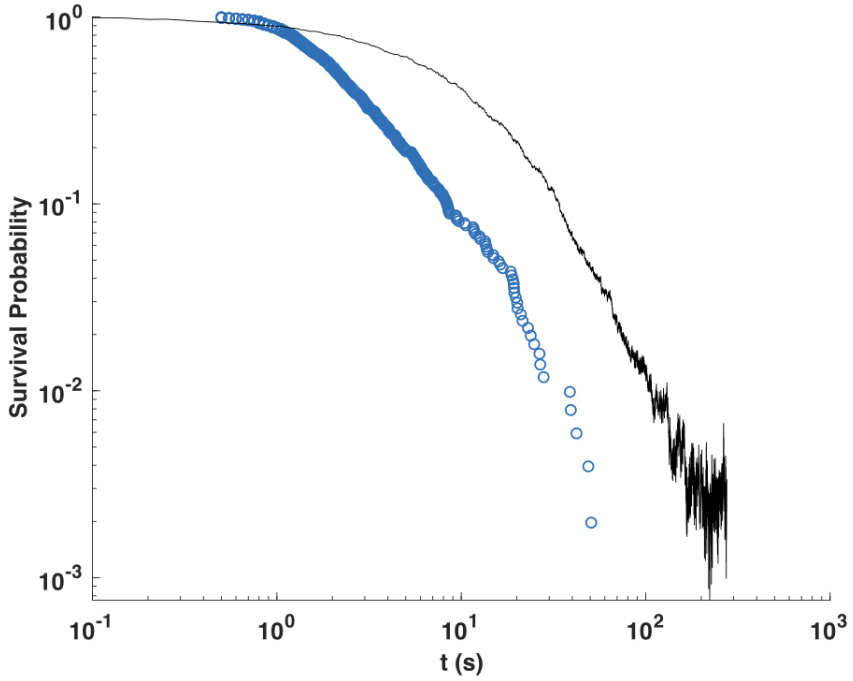



***Figure 2.13: Survival probability and bleaching curve**. Example of Survival Probability distribution of Cohesin (blue dots) and of the bleaching curve of the Halo-JF549 organic dye (black curve). The two curves are issued from the same imaging conditions (exposure time, 50ms, and laser power).*

The duration of each trajectory is considered as its binding time. From the inverse cumulative distribution of the residence times I computed the Survival Probability, defined as follow:

$$SurvivalProbability = \int_{t_0}^{\infty} P(t)dt \tag{2.4}$$

or the probability for a molecule to stay bound longer than t0. In eq. 2.4, $P(t)$ is the empiric distribution probability of the proteins residence times.

When we want to quantify the residence time of a protein with fluorescence, we are intrinsically limited by the fluorophore bleaching, as shown in

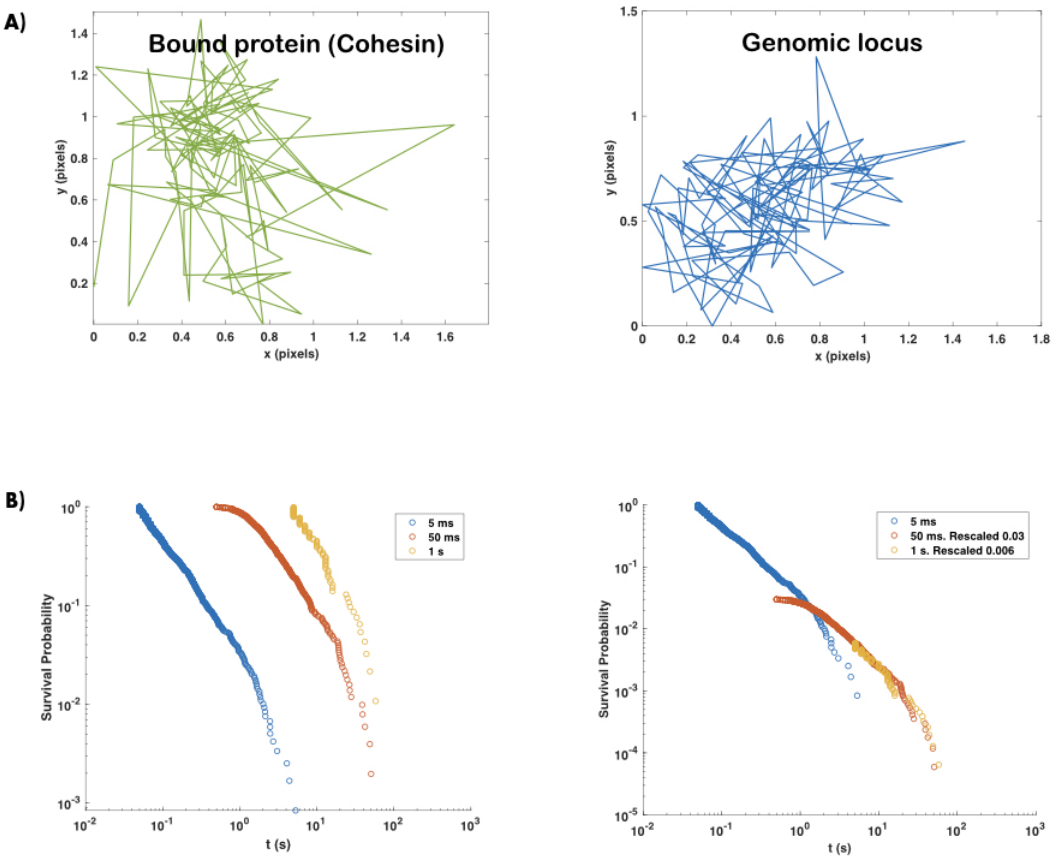



***Figure 2.14: Bound molecules and Survival probability at different timescales.*** ***A*** *Example of the trajectory of a bound molecule (left) and of a genomic locus (right), on the same time scale (122 localisations at 20Hz).* ***B*** *Left: The three dotted curves correspond to continuous acquisitions with an exposure time of 5ms (blue), 50ms (orange) and 1s (yellow). Right: The same distributions rescaled.*

fig 2.13. To overcome this limit I acquired data at different frame rates, in particular to explore longer time scales I increased the exposure time and coherently lowered the laser power. In fig 2.14 I present an example of three Survival Probabilities from experiments at different acquisition rates: 197 Hz (5ms exposure time, plus a little delay of frame transfer), 20 Hz (50ms exposure) and 1 Hz (1s exposure). The second and third Survivals (50ms and 1s exposure respectively) can be rescaled by taking the value of the first

points of the 50ms distribution and multiplying it by the value of the Survival at the same time point (Normanno et al., 2015). The result for this specific example and the rescaling weights are shown in the right plot of fig 2.14.B.

The bound molecules isolated with the selection method described above show a dynamics similar to the one of genomic loci. In fig 2.14.A I show an example of the trajectory of a bound protein and a genomic locus; the two trajectories have been chosen to have the same length. The figure is meant to show that bound proteins and genomic loci have similar dynamics.

One limitation of the method used to select the bound molecules is that I exclude the transient binding events that take place within a trajectory. This choice may affect the beginning of our distribution which would result depopulated (short binding events would last fractions of seconds), but this is not dramatic as I am trying to detect long stable binding events, trying to push further the tail and not the initial plateau of the distribution.

### 2.5.3 Analysis of dynamics

The trajectories obtained from experiments at 20Hz were analysed with custom codes implemented in Matlab. First, we computed the time-averaged MSD as

$$MSD(t) = \langle [r(t) - r(0)]^2 \rangle = \int \int P(r')(r - r')^2 P(r \mid r', t) dr dr'. \quad (2.5)$$

with $r(t)$, the position at time $t$, $r(0)$ the initial position of the particle (time $t = 0$), $P(r)$ is the steady-state distribution of the particle position and $P(r \mid r', t)$ is the probability that a particle in $r'$ will be at $r$ after a time lag $t$. Following the approach secribed in (Qian et al., 1991), I compute the $MSD$ from individual trajectories by considering the time average as follows:

$$MSD = \int (\mid r(t + t') - r(t') \mid)^2 dt'.$$

In a single cell a continuous distribution of sub-diffusion and brownian motion is often observed, an example of MSDs of CTCF molecules is shown in fig 2.15.

Once computed the MSD we extrapolated the *apparent* diffusion coefficient ($D$) [4] from each trajectory by fitting the MSD from point 2 to point 6. In this sense the apparent diffusion coefficient extracted can be thought as

[4] $D$ is called apparent since some trajectories show transitions in diffusing regime; such behaviour is the result of a bound molecule that unbinds and start diffusing or *viceversa*. Since the MSD computation goes through an time-averaging of the whole trajectory, such

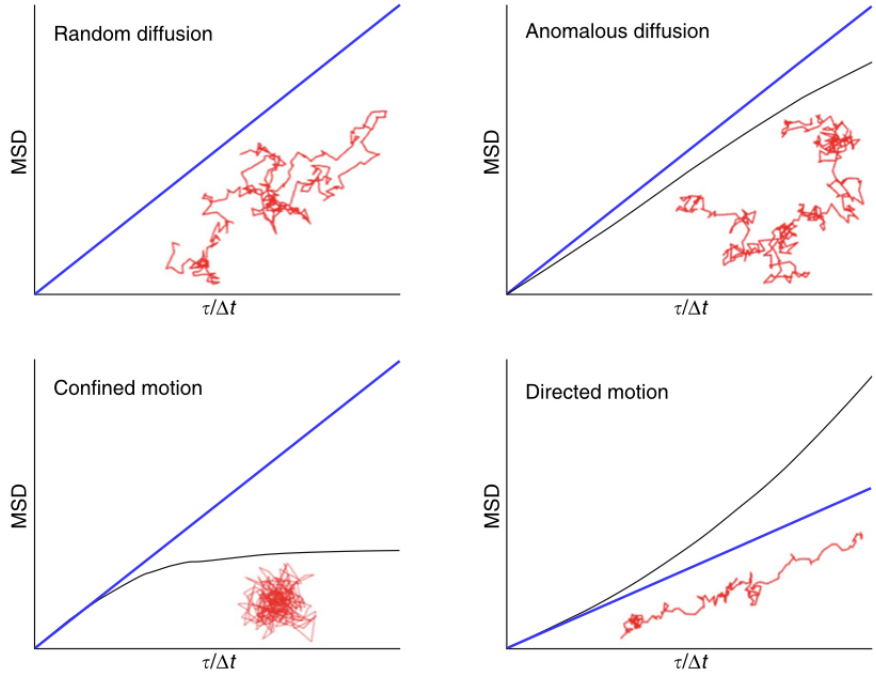



***Figure 2.15: Examples of MSD for different physical contexts****. The MSD reveals how the tracer explore the space. (Adapted from (Bradshaw and Stahl, 2016), High speed localization microscopy, El Beheiry and Dahan, pages 121-128)*

the diffusion constant at a specific timescale, in particular at $150ms$ as the timelag is $50ms$, hence $D = D_{150ms}$. For this purpose, MSDs were computed for trajectories with at least 10 localisations. To extract a diffusion coefficient I assume a purely random diffusion, also known as Brownian motion. In this scenario the MSD scales linearly with time:

$$MSD = 2nDt$$

where n is the spacial dimension, in our case n = 2 since I performed the imaging in 2D.

This diffusive phenotype correspond to the top-left panel in fig 2.15. The assumption is fair when considering the very beginning of the MSD, as shown

effect is smoothed and distributed along the entire MSD curve. As a consequence $D$ does not necessarily reflect the instantaneous mobility but a temporal average.

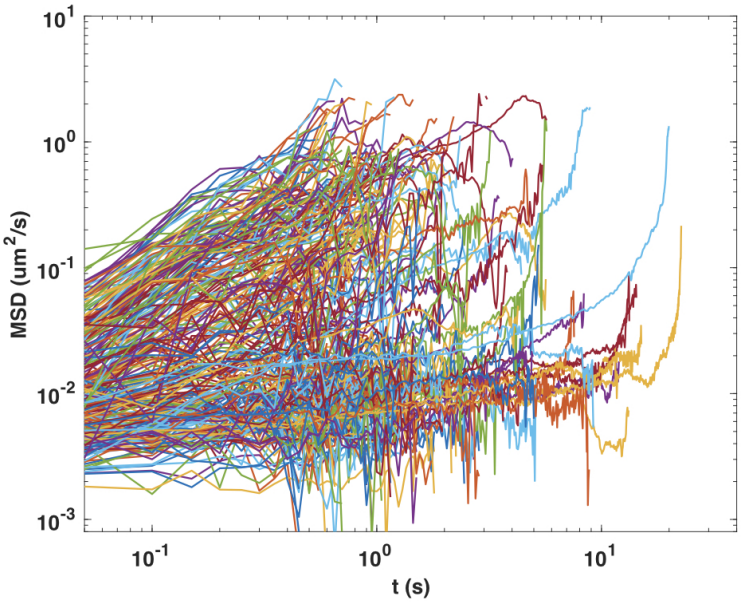



***Figure 2.16: Experimental MSD of CTCF.*** *MSDs of CTCF computed from experiments at 20Hz acquisition rate. N = 2 cells; n = 288 trajectories.*

in (Saxton, 1997). In its most general formulation the MSD does not always depend linearly on time:

$$MSD = At^{\alpha}. \tag{2.6}$$

$\alpha$, the so-called *anomalous exponent*, can be an indicator of the physics behind the detected dynamics. When $\alpha \neq 1$ the prefactor $A$ does not have the meaning of a diffusion coefficient (indeed its physical dimensions are not a $\frac{\text{space}^2}{\text{time}}$ but rather $\frac{\text{space}^2}{\text{time}^{\alpha}}$). As shown in fig 2.15, $\alpha \leq 1$ is the symptom of anomalous diffusion, also known as sub-diffusion (Top-Right), while $\alpha \geq 1$ is an evidence of super-diffusion, which correspond in some cases to directed motion (Bottom-right). I will not comment on confined motion (Bottom-left), a non trivial problem that can't be described with the simple relations mentioned above and that is not an object of study in this work. Active factors like Myosin display directed motion when walking on actin filaments, but when studying passive tracers the most common scenarios are pure diffusion and sub-diffusion. Sub-diffusion is a complex phenomenon whose phenotype can arise from very different physical scenarios. Sub-diffusion can testify the presence of many energetic traps in which the protein can fall, it can also be a descriptor of the environment and physical obstacles. These are actually two of the many reasons behind anomalous sub-diffusion and I do not even

dare to go deeper as the topic can be the object of a PhD thesis on its own. I want to mention that in the work here presented, I limited the analysis to a basic computation of the anomalous exponent to determine weather the proteins were performing pure brownian of sub-diffusive motion. To do so I took the $log_{10}$ of expression 2.6

$$log_{10}(MSD) = log_{10}(A) + \alpha log_{10}(t) \tag{2.7}$$

and I perform a linear fit to extract $\alpha$. [5]

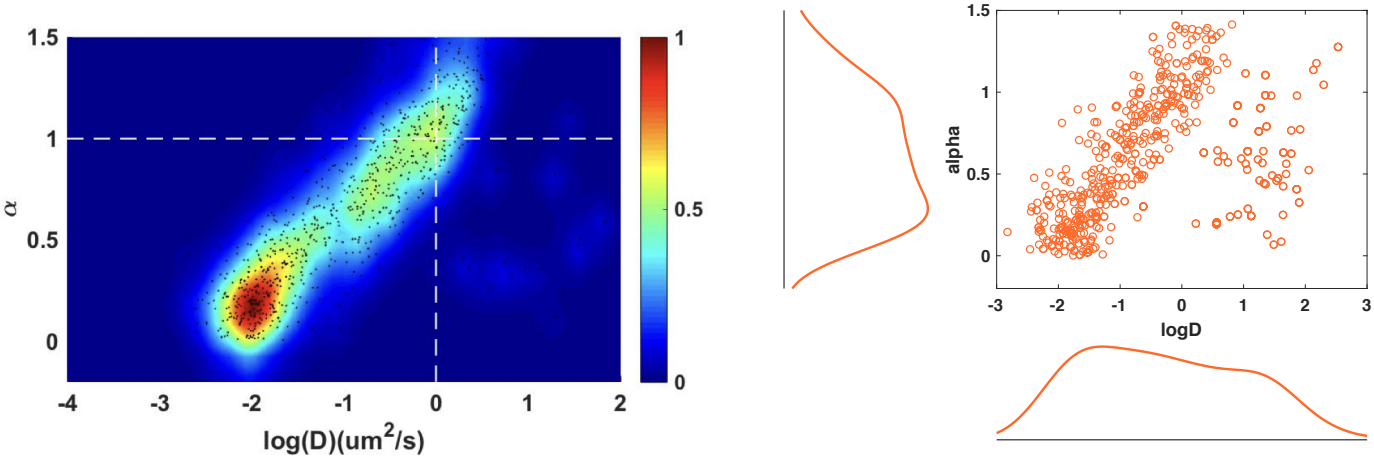



***Figure 2.17: Heatmap of $\alpha$ vs logD.*** *(Left) Heat map of the anomalous exponent $\alpha$ vs logD per trajectory. The colormap indicates the local density of points. (Right) The same data shown without the colour map for density. From tracking data on CTCF-Halo cell line, imaging performed at 50ms exposure time. Trajectories selected for this analysis consist of at least 10 localisations. N = 9 cells; n = 915 trajectories.*

Following the approach described in (Etoc et al., 2018), I plotted heat maps of $\alpha$ vs the diffusion coefficient ($D$), extracted from the linear fit of

[5] One could argue that this expression is lacking the localisation error but this is due to the fact that I chose to avoid a three parameters fitting. If I had to consider such error the expression would be $MSD = At^{\alpha} + B$ that implies $log_{10}(MSD) = log_{10}(At^{\alpha} + B)$, where $A$, $\alpha$ and $B$ are the three parameters to extract. In the two cases the distributions of $\alpha$ are comparable, a part from a few outliers, thus I preferred avoiding a three-parameter fitting.

the MSD assumed to be Brownian, of each trajectory. The results for CTCF are shown in fig 2.17. In the following chapter I will show the data without superimposing the density colour plot to enable the reader to appreciate the raw data as the computation of the local density can sometimes lead to misinterpretations. The only exception regarding the data obtained from the tracking of a genomic locus, as in this case the diffusive species are of more direct interpretation.

# Chapter 3

# Results and discussion

This chapter is dedicated to present the results obtained from the study of nuclear factors involved in the regulation of chromatin organisation in mouse embryonic stem cells. Since this work is based on tracking of DNA-binding proteins, a fluorescent chromatin locus was tracked as a reference of the template mobility, section 3.1. In sections 3.2 and 3.3, the results of respectively CTCF and Cohesin characterisation are presented.
In the following sections I will refer to Rad21 tracking results as Cohesin; it has been shown that the Cohesin complex is stable as Immuno Precipitation (IP) assays pull down the fundamental subunits Rad21, Smc1a and Smc3 (Hansen et al., 2017). Besides the characterisation of each of the two nuclear factors, Cohesin has been studied in the context of various mutations. A considerable part of the work concerns the effects of CTCF depletion on Cohesin dynamics, presented in section 3.4. In section 3.5 I will present the results obtained by tracking Cohesin in absence of Sororin 3.5.1 and Nipbl (Scc2) 3.5.2.
As described in section 2.5 different acquisition rates were chosen to quantify different observables. The choices are resumed in the following table:

| **Acquisition rate** | **Observable** | **minimum lenght of trajectory (frames)** |
|---|---|---|
| 197Hz | Fraction of bound molecules | 2 |
| 20Hz | Survival Probability<br>apparent Diffusion coefficient ($D$)<br>anomalous exponent $\alpha$ | 10 |

## 3.1 Chromatin as a reference

The object of this work is the characterization of the dynamics of nuclear proteins, whose role in the regulation of chromatin structure brings them to bind chromatin. As a bound molecule reflects the underlying diffusion of chromatin, it is worth reporting the dynamics of a locus to have a reference.

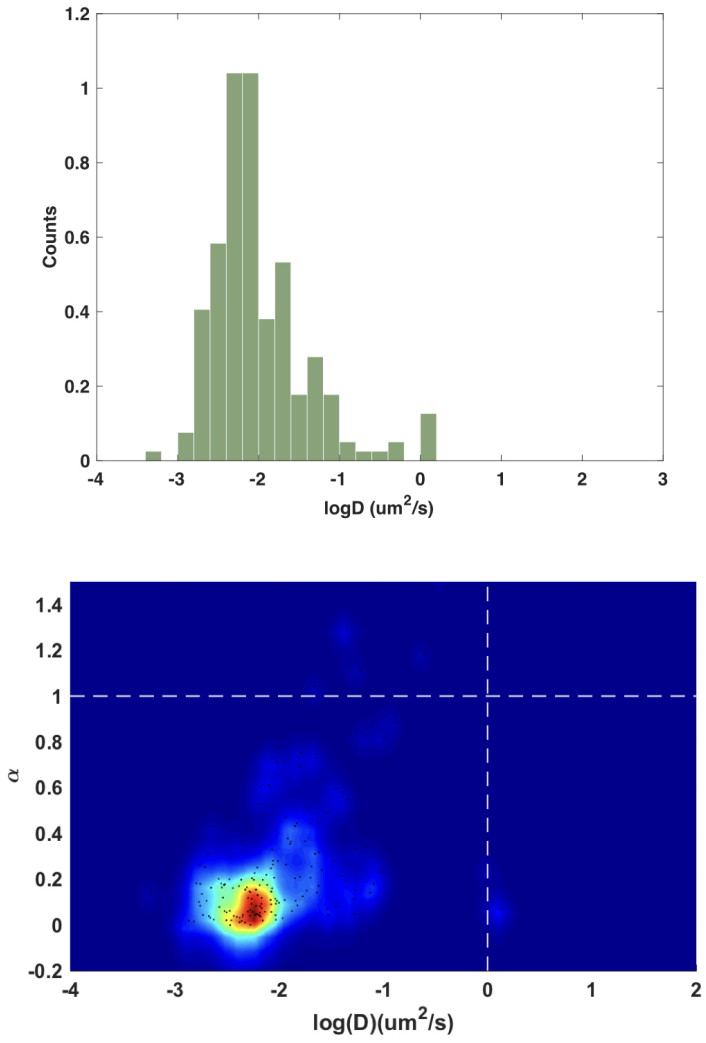



***Figure 3.1: Dynamics of XIC on the X chromosome Top*** *Histogram of diffusion coefficient of XIC.* ***Bottom*** *Heatmap of the logD (same values as the histogram) vs anomalous exponent $\alpha$. For both plots: $N = 8$ cells; $n = 176$ trajectories.*

For the sake of consistency with the cell lines used to track CTCF and Cohesin, I focused on chromatin dynamics in mouse Embryonic Stem Cells (mESC). It has indeed been shown that differentiation, or more precisely the changes in the transcriptional scenario of a tissue with respect to a polipotent cell, can heavily affect chromatin mobility (Gu et al., 2018). The experiments presented in this section has been possible thanks to the generosity of Prof. Edith Heard and her team, who kindly agreed to share the cell lines. I here present the results obtained for the X Inactivation Centre (XIC) on the X chromosome.

The cell line used to track the X inactivation centre (*Xic*), inserted via a TetO array and labelled with eGFP, has been published in (Masui et al., 2011) and (Giorgetti et al., 2016).

The values of the apparent Diffusion coefficient ($D$) are limited to a region of very small values ($-3 < logD < -1$) and the anomalous exponents obtained for these trajectories are significantly smaller than 1. For $\alpha$ the values obtained by fitting the MSD are very small, sometimes too close to 0; this effect may be due to the finite length of the MSD, as for an object that is not moving too much the first points are dominated by the localisation error. Clearly chromatin diffusion is localised in a parameter space of $\alpha < 0.5$ and $logD < -1$; these values will be used as a reference to identify and cross-validate the subpopulation of DNA-bound molecules in the following sections.

## 3.2 CTCF

To characterise CTCF binding kinetics and dynamics I tracked the protein, coupled with an Halotag, in mouse Embryonic Stem Cells. The cell line is schematically represented in panel D of 3.2: endogenous CTCF is coupled with the Halotag to perform single molecule tracking.

The different observables, i.e. the fraction of bound molecules, the Survival Probability, the distribution of diffusion coefficients, were extracted with the methods described in section 2.5 and all the results are summarised in fig 3.2. In panel A the fraction of bound CTCF and a control are reported. Half of CTCF molecules are bound in S/G2, a significantly higher value than the control. As already mentioned the fraction of bound molecules is estimated from the highest rate acquisitions (5 ms exposure at 197 Hz) to capture the fastest proteins.
For what concerns the residence time distribution, CTCF shows a power law trend (panel B). Our hypothesis is that such distributions arise from the con-

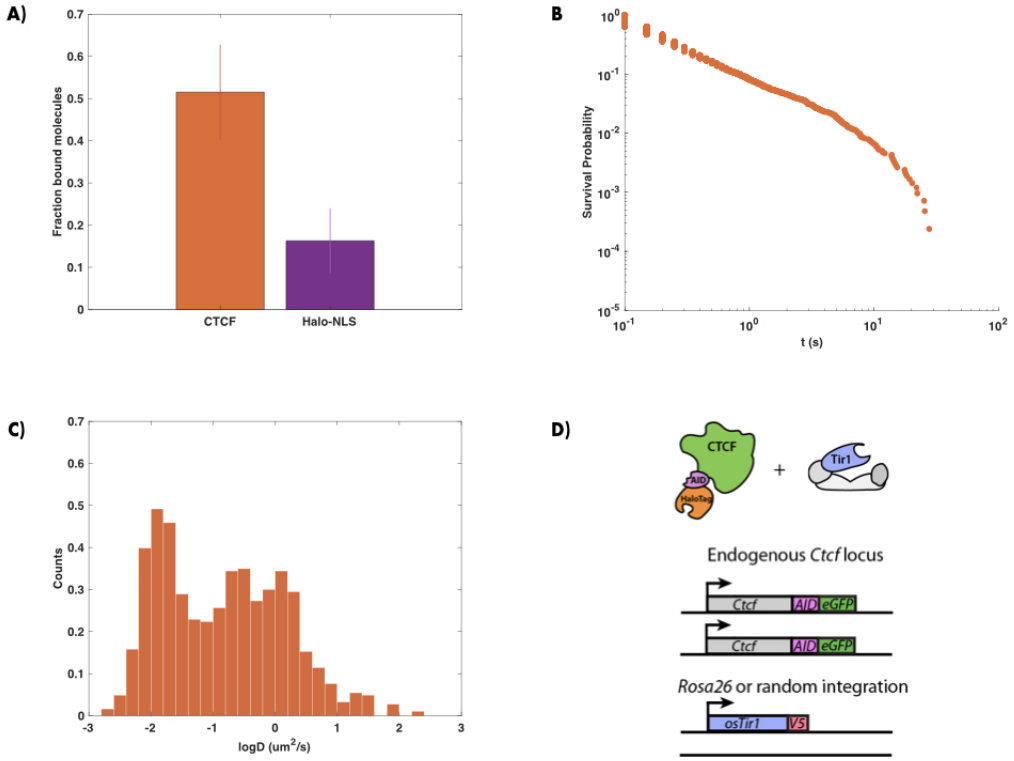



***Figure 3.2: Results from single molecule imaging of CTCF.*** ***A*** *Fraction of CTCF bound molecules and control, from 5ms exposure experiments. CTCF, N = 13 cells, n = 5344 trajectories; control, N = 10 cells, n = 2247 trajectories.* ***B*** *The Survival Probability distribution for CTCF imaged at 50ms exposure. N = 153 cells, n = 4175 trajectories.* ***C*** *Histogram of the log10(D), where D is the apparent diffusion coefficient extracted from the MSD. Bin size = 0.2; N = 13 cells; n = 915 trajectories* ***D*** *Sketch of the mESC cell line fot CTCF imaging. CTCF is coupled to an Halotag for tracking, to the Degron responsive domain for its depletion and is labelled with GFP. CTCF sequence was edited on both alleles at the endogenous site. The Tir1 box, was randomly inserted in the genome.*

volution of many dissociation rates related to transient binding events. It is indeed a recurrent behaviour of very different transcription factors searching for their target (Lac repressor: (Caccianini et al., 2015), Tet repressor: (Normanno et al., 2015) and unpublished data on Transcription Activator Like Effectors (TALE)). In the range of the fraction of seconds we are most likely

sampling non-specific interactions. The longer binding events, of the order of hundreds of seconds, are thought to correspond to stable binding events, most probably at the specific target sequence.
To describe CTCF mobility I computed the MSD for each trajectory and extrapolated the apparent diffusion coefficient $D$, the histogram of $log10(D)$ is presented in panel C of the fig 3.2. The distribution spans over a wide range of values, from highly mobile ones (between 1 and 10 $\mu m^2/s$) down to values of D that correspond to chromatin diffusion ($D \leq 0.01 \mu m^2/s$). Two populations are visible in the histogram, corresponding to a subpopulation of bound molecules (centered around $logD \sim -2$) and one of diffusing ones (centred at $logD \sim 0$). But, for the arguments exposed above concerning the number of potential dynamic subpopulations I prefer to present the raw histogram. A mean to estimate the bound fraction of molecules is to look at how many values of $D$ fall below the dynamic threshold imposed by chromatin diffusion that is $logD \leq -1$ (Gu et al., 2018) and personal work (see section 3.1). Out of 915 trajectories, 424 show a diffusion coefficient comparable to the one of chromatin, consistently with the value found by looking at the distribution of step lengths (panel A).

Concomitant to my work, several studies of CTCF single molecule tracking were published; for this reason I decided to focus my investigation on Cohesin and did not push further the analysis of CTCF behaviour. The two works (Hansen et al., 2017), (Agarwal et al., 2017) will be presented in comparison to what I did in section 3.6.

## 3.3 Cohesin

Following the approach described in section 2.5 I quantified Rad21 dynamics in Wild Type (WT) conditions in the cell line whose genotype is shown in fig 3.3.

As shown in fig 3.5 panel A, Cohesin bound fraction is around 70%, significantly higher than the value found for CTCF. Stem cells spend more than 60% of the cycle in S phase, as shown by (El-Badawy and El-Badri, 2016), and the Hoechst staining enabled us to confirm that cells were not undergoing Mitosis. We assume the value observed for the fraction of bound molecules to correspond to cells in S/G2, thus the result is consistent with what published in (Gerlich et al., 2006).

The Survival probability shows a power law decay and reaches higher values for the longest binding events. It is interesting to observe that a protein that does not have a specific DNA binding domain shows the same

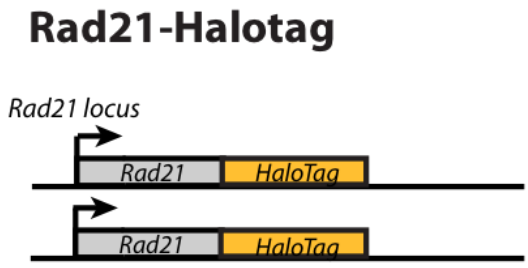



***Figure 3.3: Genotype of the cell line used to track Cohesin.***

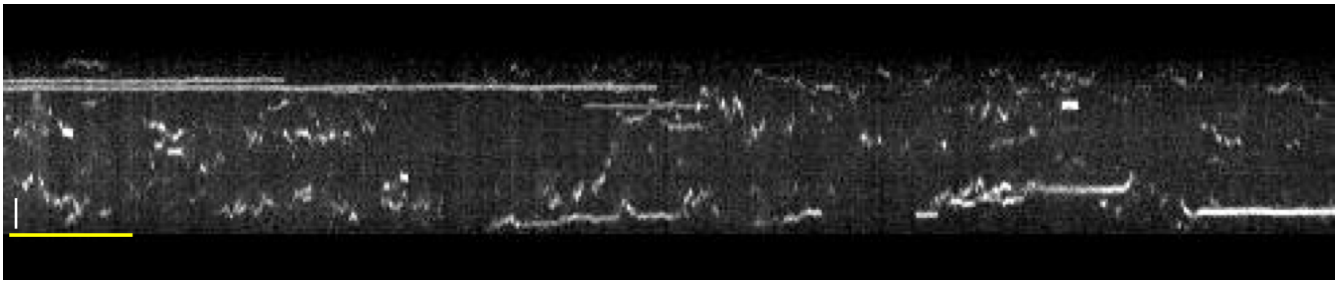


***Figure 3.4: Kymograph of Rad21.*** *White vertical line: spatial scale-bar* $2\mu m$*. Yellow horizontal line: time scale-bar 5 s.*

behaviour as transcription factor-like proteins. This tells us that Cohesin performs plenty of non-specific interactions with DNA [1].

Interestingly the tail of the distribution reaches higher residence times than CTCF, in particular it decays around 60s, right before bleaching. It is has been shown via FRAP experiments that Cohesin stable bindings can last up to 20-30 minutes (Gerlich et al., 2006) (Ladurner et al., 2014) (Hansen et al., 2017), but with single molecule tracking is difficult to reach such timescales. The plot presented in fig 3.5 panel B is issued from the same experimental conditions mentioned for CTCF in the previous section ($t_{exp} = 50ms$ continuous imaging). In panel C the histogram of $logD$ is shown. The distribution is very broad, representing a very heterogeneous diffusive sce-

[1] Throughout the manuscript I will refer to non target specific interaction as "non-specific interactions".

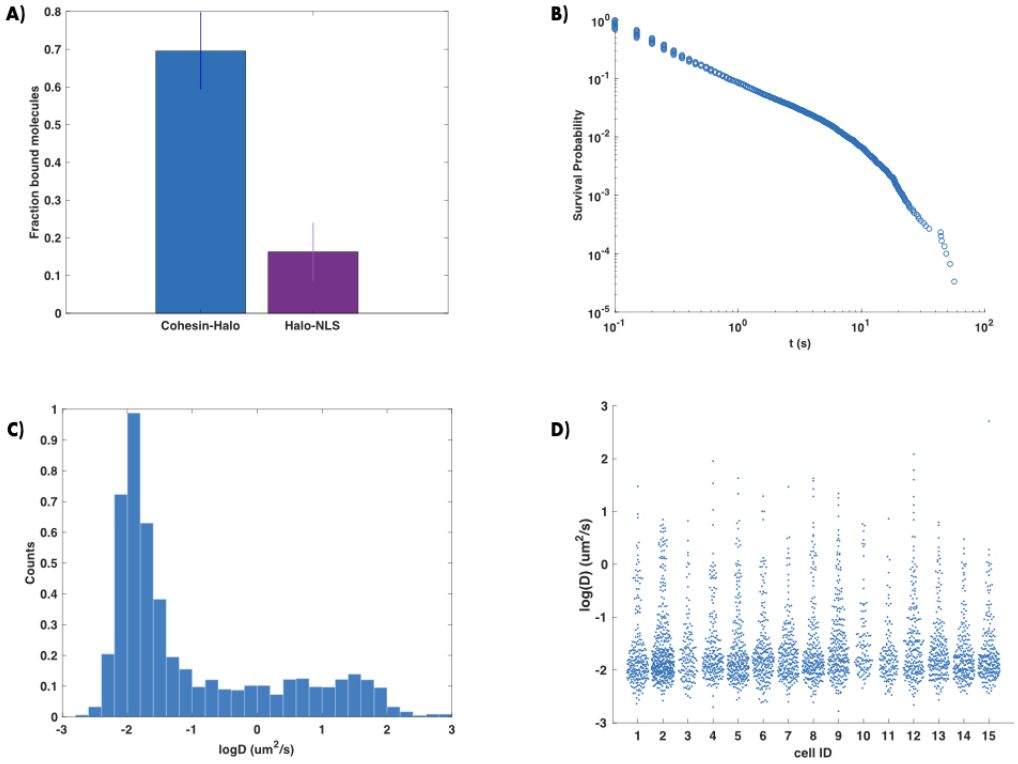



***Figure 3.5: Results from single molecule tracking of Rad21.*** ***A*** *Fraction of bound Rad21 bound and control, from 5ms exposure experiments. N = 13 cells.* ***B*** *The Survival Probability distribution for Rad21 imaged at 50ms exposure. N = 15 cells; n = 29975.* ***C*** *Histogram of the log10(D), where D is the apparent diffusion coefficient extracted from the MSD. N = 15 cells; n = 6810 trajectories.* ***D*** *Plot of the logD data showing the cell-to-cell variability.*

nario. There is a considerable peak around $logD \sim -1.75$, corresponding to bindings events, and another peak around $logD \sim 1.5$, which represents freely diffusive molecules. In between these two peaks there is a considerable intermediate region, $-1 < logD < 1$ that can't be easily associated to a specific diffusive population. Differently from the majority of nuclear factors, in particular DNA-binding proteins, Cohesin stands out for its highly heterogeneous mobility. The diversity of Cohesin dynamics can be appreciated from the kymograph in fig 3.4. In panel D I reported the cell-to-cell variability for the values of $logD$: the distributions par individual cells are very similar, hence the data reported in panel C are a robust representation

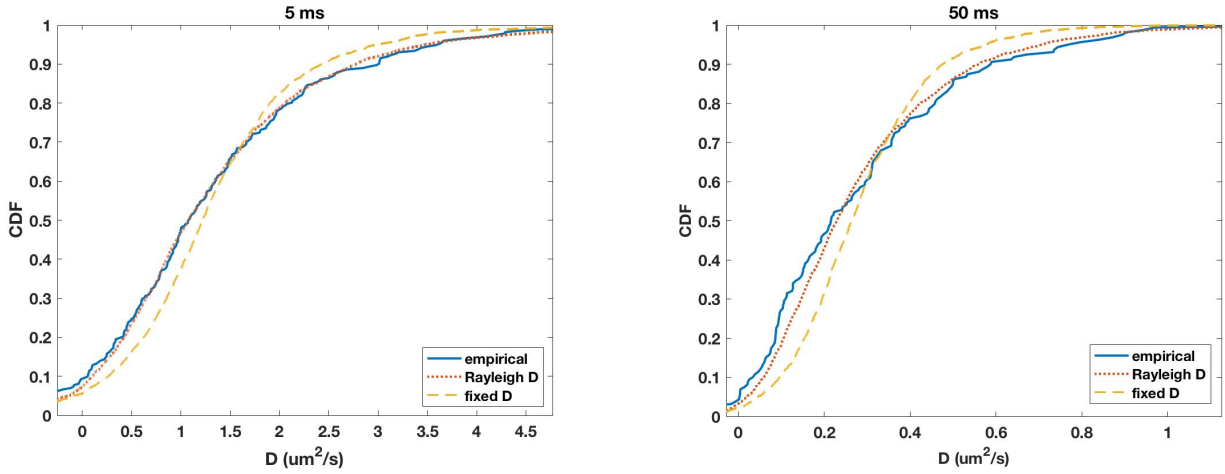



***Figure 3.6: Cohesin's diffusion cannot be modelled with one single value of the diffusion coefficient*** $D$***.*** *The distribution of the experimental diffusion coefficients* $D$ *cannot be reproduced if the trajectories are simulated based on a single value of* $D$ *as an input. The experimental results are retrieved if* $D$ *comes from a Rayleigh distribution.*

of the diffusivity scenario in S/G2 for Cohesin in mouse embryonic stem cells.

The heterogeneity of Cohesins diffusive behaviour is confirmed by simulations performed by Simon Grosse-Holz, from the group of Leonid Mirny (MIT, Cambridge, MA). The pipeline of the simulation is the following:

1. generate a random walk with diffusion coefficient $D$

2. add localisation error $\sigma$ and motion blur

3. repeat, generating an ensemble of trajectories whose lengths and acquisition times match the real data.

The values of $D$ and $\sigma$ used for the simulations are estimated from experimental data. As shown in fig 3.6, a single value of $D$ is not sufficient to reproduce the distribution observed experimentally. To match the experimental and the synthetic data it is necessary to choose not a single value of $D$ but a distribution $P(D)$, in our case a *Rayleigh* distribution.

We are currently focusing our work on the investigation of the relationship between the spread of the diffusion coefficient values and possible crowding.

## 3.4 Cohesin without CTCF

One of the goals of this work is to study the interplay between nuclear factors involved in the regulation of three-dimensional chromatin structure. To achieve such goal we decided to first study the dynamics of Cohesin in absence of CTCF. Thanks to the work of Elphège Nora, in the laboratory of Benoit Bruneau at the Gladstone Institute in San Francisco (USA), who produced the cell line illustrated in fig 2.7, I could image the Cohesin subunit Rad21 in the same conditions of his previous work, (Nora et al., 2017) in which he showed that CTCF depletion triggers loss of TADs, loss of TAD insulation, and loss of Cohesin positioning at CTCF sites. By performing single molecule tracking in living cells I could address the problem from a dynamic point of view and keeping the information of individual molecules, in single cells. The genotype of the cell line is sketched in fig 3.7.
Rad21, or Scc1, is a stable component of the Cohesin complex; it has been shown by co-immuno precipitation that precipitating Rad21 implies the precipitation of all the other native components of Cohesin (Smc1a, Smc3) (Hansen et al., 2017). It is fair to assume that when imaging Rad21 we are observing the entire complex and I will eventually use Rad21 and Cohesin as synonims.

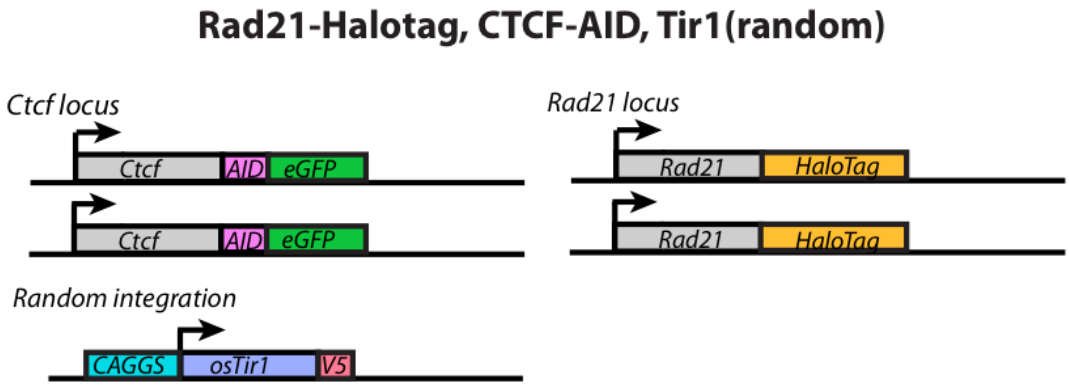



***Figure 3.7: Genotype of the cell line used to track Cohesin in presence/absence of CTCF.***

To study Cohesin in absence of CTCF cells were incubated overnight (∼ 14 hours) with auxin. Despite the kinetics of the degron system is on the scale of a few hours (see fig 2.8), we chose longer incubation times to reach homogeneity in the cell population, some kind of biological steady state. Proper deletion was checked in fluorescence thanks to the GFP reporter coupled to CTCF (see fig 3.8, central panels). To identify the nuclear boundaries, es-

pecially in absence of CTCF, I stained chromatin with $1\mu M$ of Hoechst (see fig 3.8). Experiments on treated and untreated cells were always conducted in the same day and the same imaging conditions.

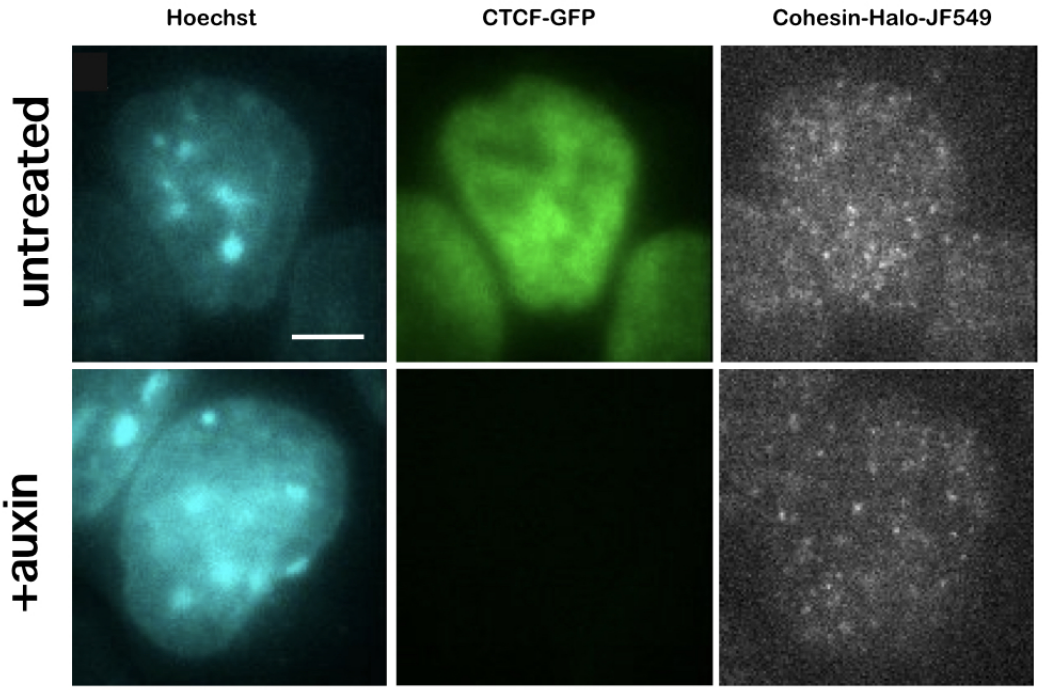



***Figure 3.8: Cohesin +/- CTCF****. The top raw correspond to untreated cells while the bottom row represents cells incubated with auxin overnight. From left to right, the nuclei stained with Hoechst 1uM, the signal in GFP indicating the presence/absence of CTCF, an image in the channel of the dye coupled to Cohesin JF549. Scale bar* $= 5\mu m$*.*

.

Coherently with the tracking strategy adopted for CTCF and Cohesin WT, I performed fast acquisition rate imaging and extrapolated, from the relative datasets, the fraction of bound molecules. The results are shown in fig 3.9 panel A, untreated cells are represented in blue and cells supplemented with auxin are in purple (the colour code is maintained throughout the whole manuscript). There is no significant difference in the fraction of bound Cohesins when depleting CTCF and the values reported are consis-

tent with what found for Cohesin WT and with the literature (Gerlich et al., 2006) (Hansen et al., 2017) for cells in S/G2. In panel B of the same figure I reported the Survival Probability for Cohesin originated from two sets of acquisitions, with $t_{exp} = 50ms$ and $t_{exp} = 1s$ both realised in a continuous imaging regime. In such imaging conditions I could acquire movies that lasted 5000 frames which correspond to $\sim$ 4 minutes. The curve derived from experiments at the longer timescale was rescaled with the method illustrated in fig 2.14 and explained in section 2.5.2, visibly the junction point of the two curves lies around 5 seconds. In the same plot the bleaching curve is displayed, it represents the decay of the dye for $t_{exp} = 50ms$ continuous imaging (with the same imaging conditions of the data acquisition). The distribution of residence times of Cohesin does not seem to be affected by CTCF depletion up to the minute time-scale. As already mentioned, it has been shown by different groups that Cohesin can stay bound to chromatin up to $\sim 20 - 30$ minutes and clearly I am not reaching this time-scale with my single molecule experiments. Consequently, I decided not to infer dissociation rates from the Survival Probabilities distributions as the resulting values would be describing only the transiently binding sub-populations of Cohesin. I will discuss the subject more in detail in section 3.6 but globally the absence of CTCF does not seem to affect Cohesin binding kinetics.

On the other hand Cohesin dynamics appear slightly different with and without CTCF. In panel C and D of fig 3.9 the distribution of the diffusion coefficient are reported. The histograms of Cohesin's $logD$ +/- CTCF have the same wide distribution, spanning over more than 4 orders of magnitude, indicating that there is no effect on a particular diffusive sub-population of Cohesin. An increase of the mobile molecules is visible in the histogram and becomes more striking when computing the Cumulative Distribution Function (CDF), shown in panel D. CTCF depletion implies a significant increase of the mobile sub-population. The discrepancy between the two CDFs was quantified via a Kolmogorov-Smirnov test that returned a p-value of $p = 0.0196$, assuming that the two experimental curves arise from the same distribution. Interestingly the discrepancy between the two CDFs is localised in a range of values of $logD$ that covers three order of magnitudes ($-1 < logD < 1$) and it is the highest in a window of $logD$ that is too low for freely diffusing proteins and too high for bound molecules. To further investigate the fraction of Cohesins affected by CTCF depletion I isolated all the trajectories lying in the region $-1.5 < logD < 0$[2] and I plotted the distribution of step length, results are shown in the top of fig 3.10. There is

[2]I chose to exclude trajectories with $logD > 0$ because these are the values of freely diffusing proteins.

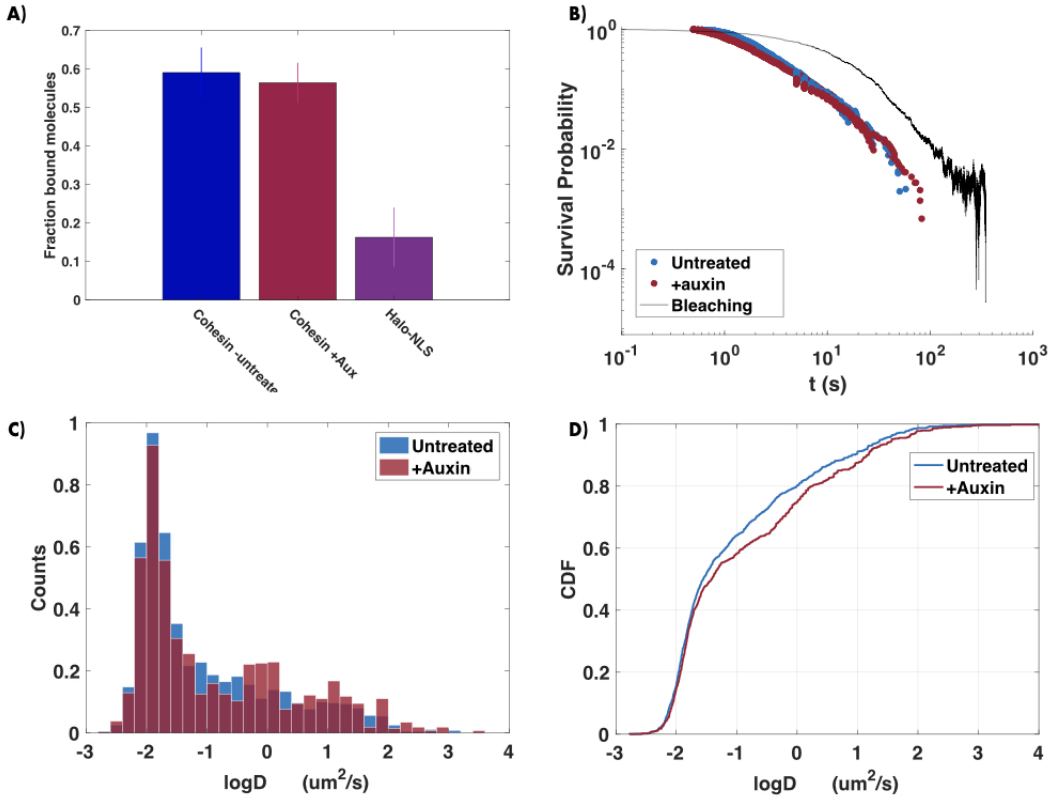



***Figure 3.9: Cohesin +/- CTCF Results.A*** *Fraction of bound Cohesin molecules in presence (blue) and absence (+Aux, purple) of CTCF and control. Untreated: N = 14 cells. +auxin: N = 16 cells.* ***B*** *Survival probability of Cohesin +/- CTCF. Untreated: N = 15 cells; n = 600 trajectories. +auxin: N = 15 cells; n = 949.* ***C*** *Histrogram of logD of Cohesin in presence and absence of CTCF. Untreated: N = 15 cells; n = 1756 trajectories. +auxin: N = 15 cells; n = 1277 trajectories.* ***D*** *Cumulative distribution function of the data presented in the histogram of panel C. This representation allows a better understanding of the differences in mobility between the two mutants.*

.

no striking difference between the two distributions and, kinetically speaking, one single population is observed in both cases. To complete the investigation on Cohesin dynamics with and without CTCF I extracted the anomalous exponent $\alpha$ from each trajectory, as explained in 2.5.3 and plot it as a function of the $D$ of the same trajectory in the scatterplots shown in fig 3.10 . A

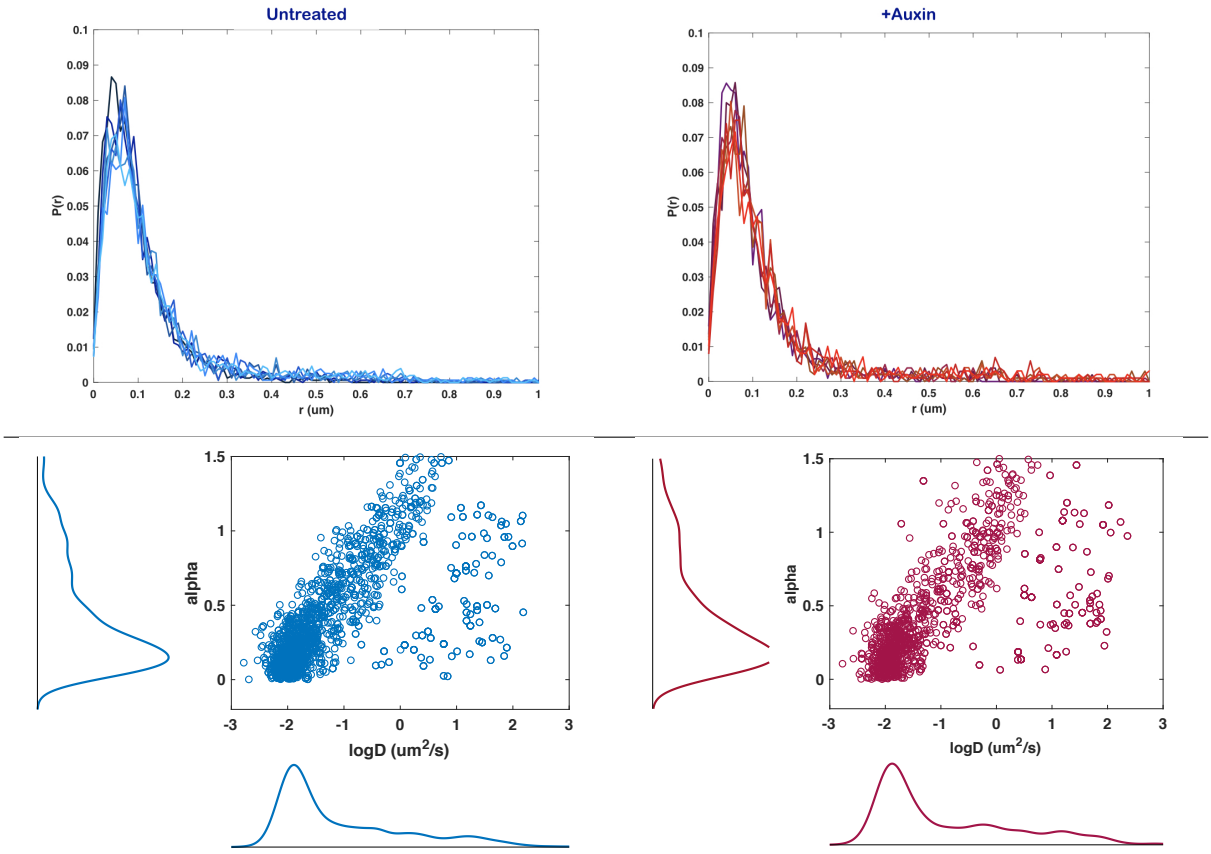



***Figure 3.10: Characterising Cohesin dynamics +/- CTCF Top*** *Distributions of step size for Cohesin in presence (Blue) and absence (Red) of CTCF for selected trajectories in the interval* $-1 < logD < 0$. *Both cases* $N = 15$ *cells. Untreated:* $n = 269$ *trajectories. +auxin:* $n = 300$ *trajectories.* ***Bottom*** *Scatter histograms of the anomalous exponent* $\alpha$ *vs the apparent diffusion coefficient* $D$. *The two parameters where extracted following the procedure described in 2.5.3. Untreated:* $N = 15$ *cells;* $n = 1756$ *trajectories. +auxin:* $N = 15$ *cells;* $n = 1277$ *trajectories. Selected trajectories with* $> 10 frames$.

very close look is needed to appreciate the mild differences between the two plots: there are less immobile Cohesins in absence of CTCF. Overall the binding kinetics does not seem to be affected by the absence of CTCF as the fraction of bound molecules stays unchanged and the residence time distribution is constant up to the minute time-scale. These two conclusions are consistent with the ChIP-Seq results shown in fig 3.27: the amount of

Cohesin on chromatin is not altered in absence of CTCF. Cohesin's dynamics is mildly affected by CTCF knock-out, but the nature of the perturbation is still unclear. There is no significant variation in the dynamic species observed for Cohesin.

### 3.4.1 Cohesin in non-cycling cells

Cohesin behaviour is tightly linked to the cell cycle, as briefly described in section 1.3.2 and described in fig 3.11. In particular in early S phase a subpopulation of Cohesins is acetylated and locked on chromatin to grant proper sister chromatid cohesion (Peters et al., 2008).

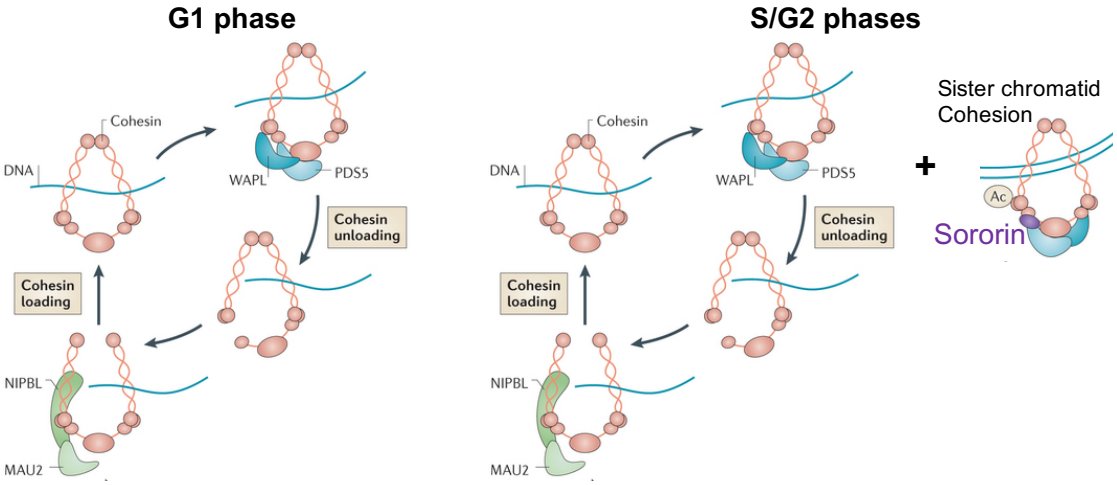



***Figure 3.11: Cohesin throughout the cell cycle.*** *Adapted from (Losada, 2014)*

This study is focused on the behaviour of Cohesin in the regulation of spatial organisation of interphase chromosomes, thus a way to discern *cohesive* Cohesins was needed, where *cohesive* Cohesin indicates Cohesin complex involved in sister chromatid cohesion, acetylated and stabilised by Sororin. Following the approach adopted by Elphège Nora in his latest work I performed the tracking experiments in non cycling cells.

Other works relied on cycle markers like PCNA (Gerlich et al., 2006) the Fucci system (Hansen et al., 2017) but such approaches do not exclude the *cohesive* Cohesins from the picture. Moreover, in the cited studies, the authors performed Fluorescence Recovery After Photobleaching (FRAP) experiments on Cohesin whose outputs are mean values on the mobility of the entire ensemble of molecules.

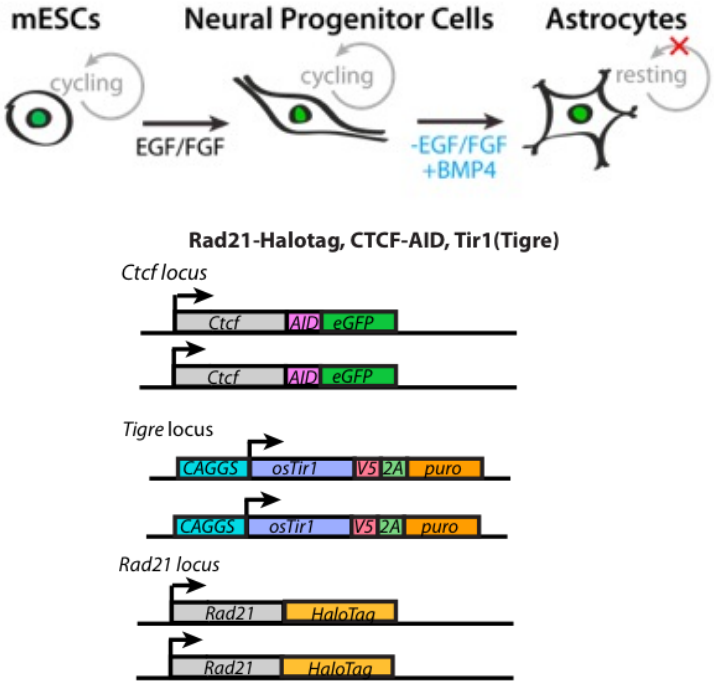



***Figure 3.12: Path to obtain resting astrocytes from mouse embryonic stem cells and genotype. Top*** *sketch of non cycling Astrocytes, adapted from (Nora et al., 2017).* ***Bottom*** *Genotype of the cell line.*

Choosing non-cycling cells gave us the certainty that each Cohesin we were looking at via single molecule tracking was not involved in sister chromatid cohesion. Consistent with published observations in (Nora et al., 2017), we noticed that the CTCF-AID, Rad21-Halotag cells stopped responding to auxin upon differentiation, presumably because of silencing of the randomly integrated Tir1 transgene. We overcame this issue by creating another CTCF-AID cell line with the Tir1 transgene targeted at the Tigre locus, which remained stable upon cell differentiation into Neural Progenitors and Astrocytes, as schematically described in fig 3.12.

As with CTCF and Cohesin I performed single molecule tracking experiments with HiLO continuous imaging at $t_{exp} = 5ms$, to quantify the fraction of bound molecules, at $t_{exp} = 50ms$, to characterise the dynamics at to extract the distribution of residence times. As indicated earlier, the Astrocytes are differentiated from the stem cells used to study Cohesin in presence and absence of CTCF, this allowed me to investigate the behaviour of Cohesin

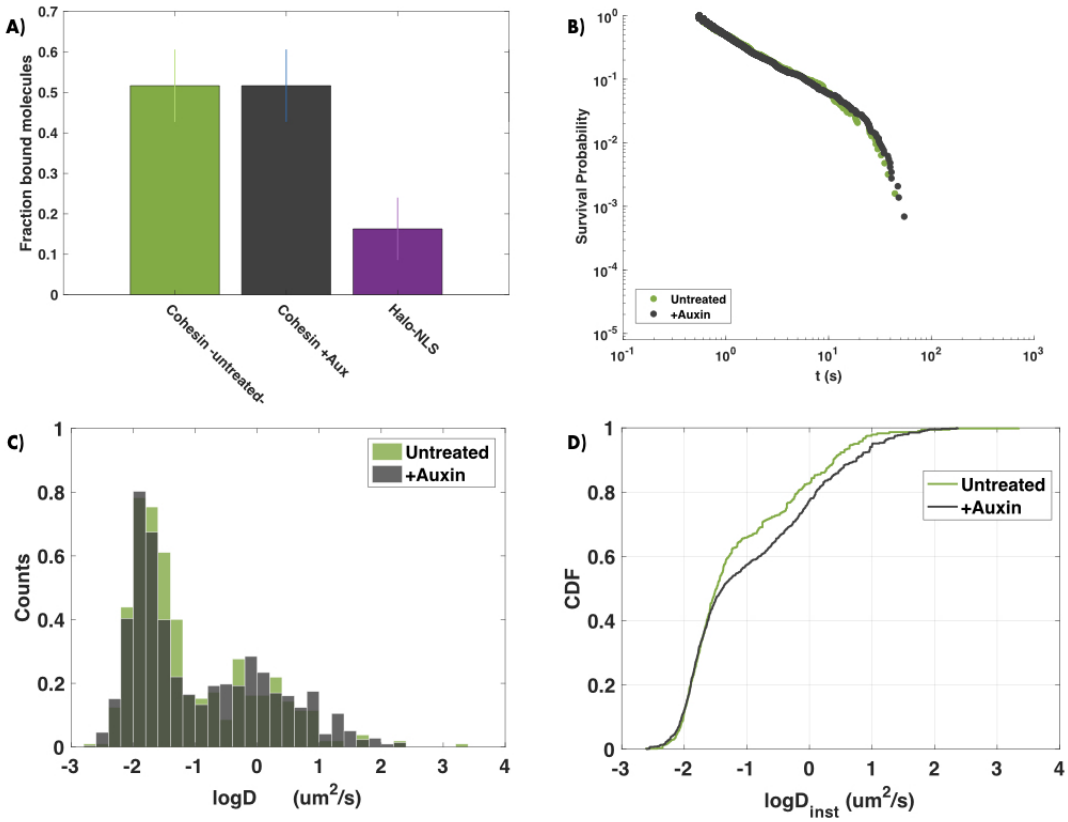



***Figure 3.13: Results of single molecule tracking of Cohesin +/- CTCF in resting Astrocytes.*** ***A*** *Fraction of bound Cohesins in G0. Clearly, there is no difference in presence and absence of CTCF. N = 15 cells in both cell lines.* ***B*** *Survival Probability obtained from tracking at* $t_{exp} = 50ms$*, N = 15 cells in both cell lines. Untreated: n = 5568 trajectories; +auxin: n = 1449 trajectories.* ***C*** *Untreated: N = 15 cells; n = 524 trajectories. +auxin: N = 14 cells; n = 1090 trajectories.* ***D*** *Cumulative Distribution Function of the data shown in the histogram in panel C. For both C and D, selected trajectories with* $> 10 frames$

+/- CTCF in G0, meaning in absence of *cohesive* Cohesins. Results are shown in fig 3.13.

50 % of Cohesins are bound in G0 and there is no difference in presence or absence of CTCF(fig 3.13 panel A). The distribution of residence times, exhibited in panel B as the Survival Probability, follow the same decay as the previously presented curves for Cohesin (results are merged in fig 3.28). Importantly in this case the difference between the two CDFs of the $logD$ is even more significant than in the ES case. A Kolmogorov-Smirnov test gave

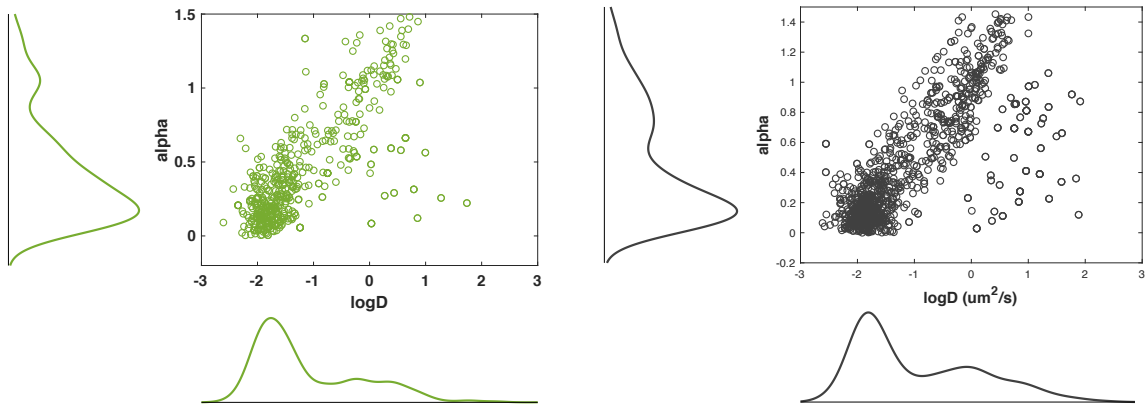



***Figure 3.14: Scatter plot of*** $\alpha$ ***vs*** *logD* ***for Cohesin in Astrocytes.*** *Untreated: N = 15 cells; n = 524 trajectories. +auxin: N = 14 cells; n = 1090 trajectories.*

a p-value of $p = 0.0014$, the null hypothesis being that the two curves belong to the same distribution. This mild difference can be appreciated also in the scatterplot shown in fig 3.14: when treating the cells with auxin we observe a slight increase of the mobile fraction ($\alpha \sim 1$ and $logD \sim 0$).

We next sought to corroborate the results in non cycling cells with experiments abrogating sister chromatid Cohesin in ES cells.

## 3.5 Other mutants

### 3.5.1 Cohesin in absence of Sororin

In this section I will present the outcome of tracking experiments on Cohesin in absence of Sororin. Before entering the details of the results I will briefly present Sororin and its role in relation to the Cohesin complex.

Sororin is a vertebrate protein required to lock Cohesin on chromatin and ensure sister chromatid cohesion. Sororin stabilises Cohesin-DNA binding by hiding the domain recognised by Wapl, Cohesin unloader. In fact, acetylated Cohesin dissociation can be achieved once Sororin is phosphorylated and unloaded from the Cohesin complex, leaving room for Wapl to open the ring and release Cohesin (Nishiyama et al., 2013). Sororin is recruited on chromatin already in S phase (Nishiyama et al., 2010) and shortly after

DNA replication (Lafont et al., 2010) and in absence of the Cohesin complex, Sororin can't be loaded. Sororin binding to the Cohesin complex is related to Cohesin acetylation: it has been shown by two different groups that the Esco1/Esco2 acetyltransferases are necessary, but not sufficient, for Sororin-Cohesin binding (Nishiyama et al., 2010) (Lafont et al., 2010). Sororin depletion dramatically affects sister chromatids cohesion and Sororin-lacking cells end up blocked in Mitosis (Rankin, 2005), as shown in fig 3.15

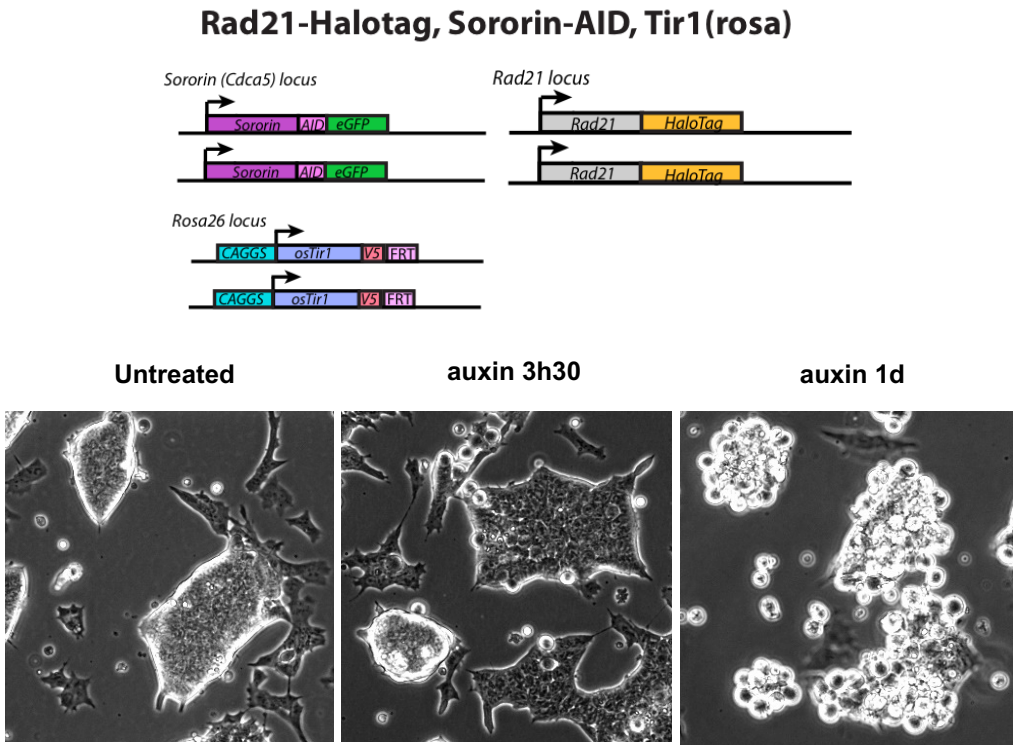



***Figure 3.15: Cell line for Cohesin tracking in absence of Sororin. Top*** *sketch of the genotype of the cell line used to track Cohesin in absence of Sororin.* ***Bottom*** *Brightfield images of cells lacking Sororin upon incubation with auxin. After 3h of auxin incubation we can already observe some cells in Mitosis and after 1 day of incubation the majority of cells are blocked in Mitosis.*

Following the depletion strategy used for CTCF (see section 3.4) Sororin was coupled to the auxin inducible degradation system in a Cohesin-Halo cell line (2.3). The kinetics of Sororin depletion upon incubation with auxin is shown in fig 3.16. I investigated the behaviour of Cohesin in absence of

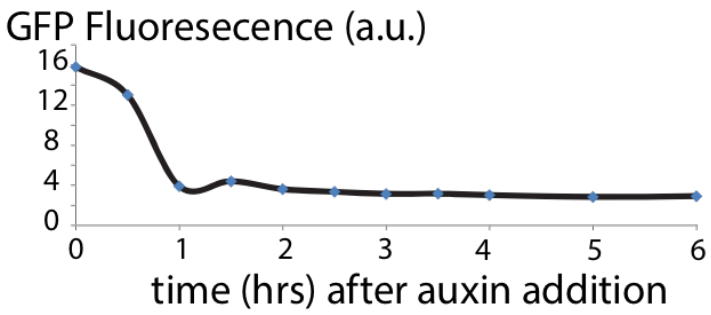



***Figure 3.16: Kinetics of Sororin degradation measured with fluocytometry.***

***Figure 3.17: Images of cells lacking Sororin after incubation with auxin for 3 hours.*** *From left to right: transmission image (a Mitotic cell is visible on top of the adherent one); image in the Hoechst channel that indicated that the cell is not yet in Mitosis; snapshot in the channel of Cohesin-Halo-JF549 from the single molecule imaging movie. Field of view = 20.5 $\mu m$.*

Sororin with two different incubation timings: after 3 hours and 6 hours of auxin incubation. Cells imaged after 3 hours incubation with auxin were mostly still in S/G2 (roughly 1 cell out of 20 had a mitotic phenotype) see fig 3.17 for an example. The chosen incubation timescale is a fine ratio between the time needed to achieve an homogeneous degradation of the protein (see fig 2.8 for the degron kinetics) while avoiding a longer incubation that inevitably induces the majority of the cells in Mitosis.

In order to exclude cells that would have blocked in mitosis early, and be able to make a direct comparison with the previous data collected in interphase, and coherently with the protocol adopted for the experiments presented in the previous chapters, cell were stained with Hoechst. In fig

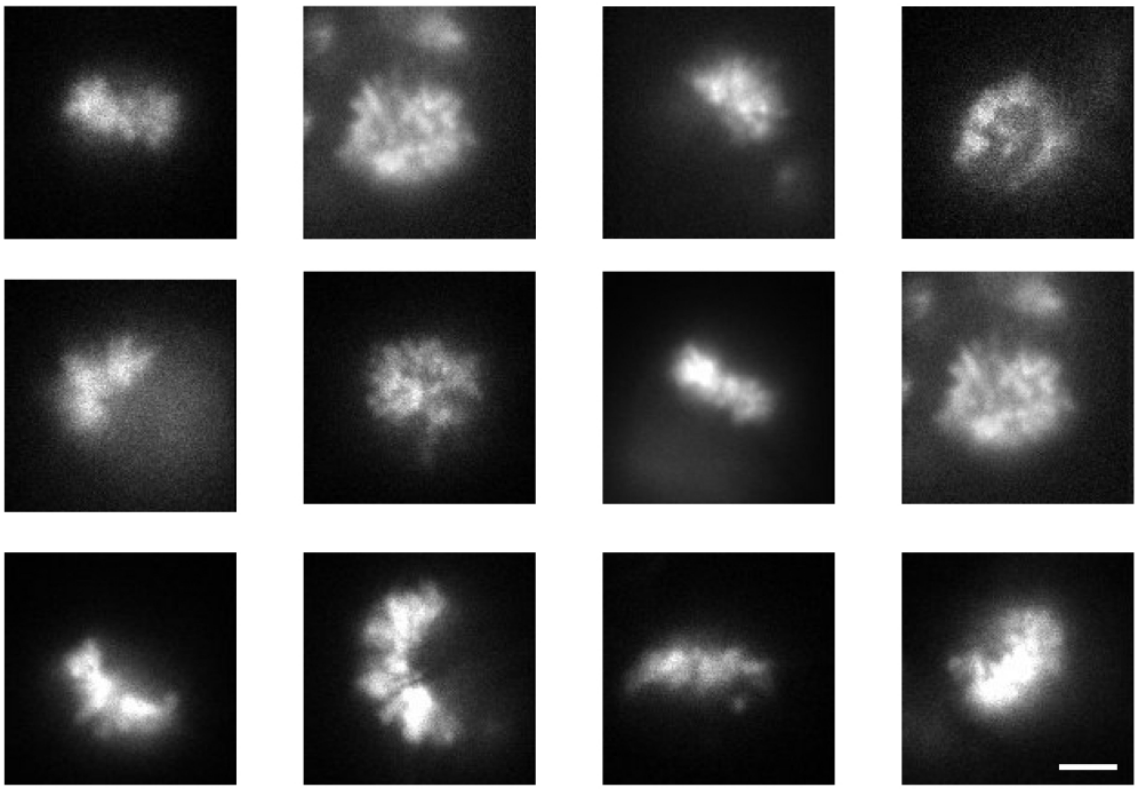


***Figure 3.18: Examples of mitotic cells after Sororin depletion for more than 6 hours.***

3.17 an example of the S/G2 imaged cells is reported, while fig 3.18 displays some of the images in the Hoechst signal in Mitotic cells.

The results of Cohesin tracking in absence of Sororin both for cells in S/G2 and in Mitosis are shown in fig 3.19. Panel A reports the quantification of the fraction of bound molecules of Cohesin in absence of Sororin in S/G2 cells (violet) and in Mitosis (emerald). The absence of Sororin mildly affect the fraction of bound Cohesins while in Mitosis the portion of bound molecules is even lower than in the control (purple). The small effect in S/G2 cells may be due to the fact that Sororin interacts with the Cohesins that are already topologically loaded onto chromatin and acetylated (Nishiyama et al., 2010) (Lafont et al., 2010) and this subpopulation of Cohesins is not very large ($\sim$ 10% from our experiments). On the other hand

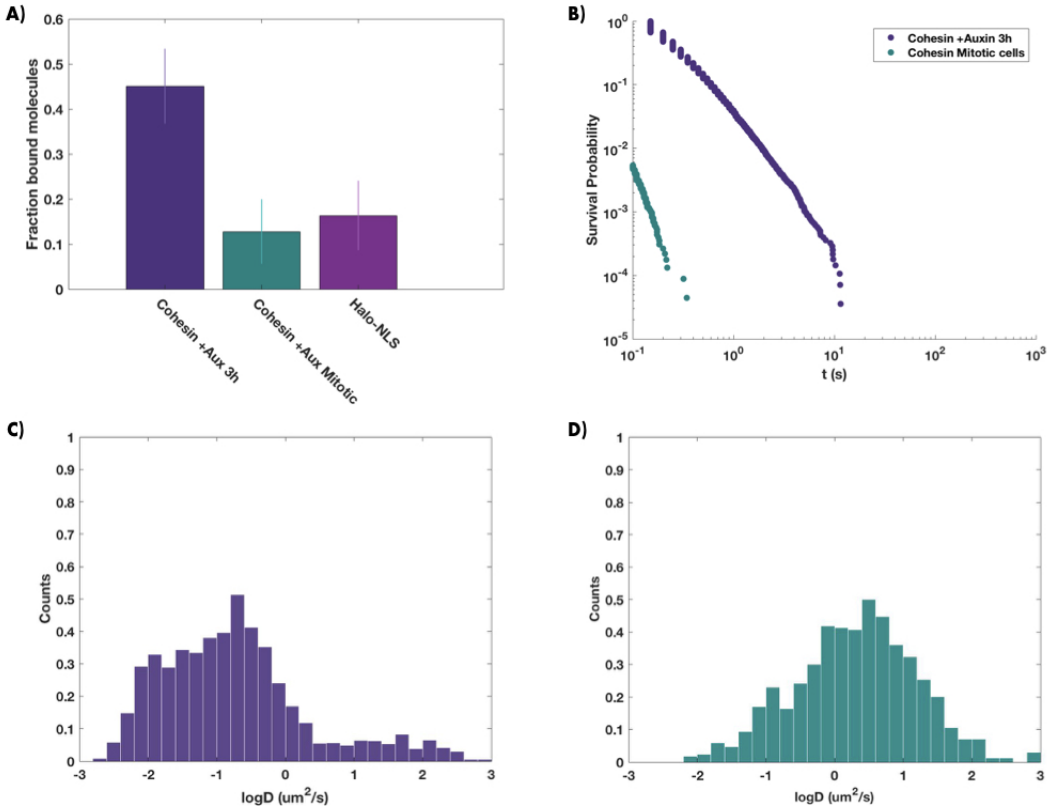



***Figure 3.19: Results of single molecule tracking of Cohesin in absence of Sororin.*** ***A*** *Fraction of bound Cohesins in absence of Sororin in S/G2 and in Mitosis. Cohesin in Mitosis can't bind chromatin anymore, even transiently, almost all molecules are freely diffusing.* ***B*** *Survival Probability obtained from tracking at* $t_{exp} = 50ms$*. S/G2:* $N = 16$ *cells;* $n = 27960$ *trajectories; Mitotic:* $N = 14$ *cells;* $n = 22729$ *trajectories.* ***C, D*** *Histograms of logD for Cohesin in absence of Sororin in S/G2 (panel C, violet) and in Mitosis (panel D, emerald). S/G2 cells:* $N = 17$ *cells ;* $n = 4021$ *trajectories. Mitotic:* $N = 18$ *cells;* $n = 849$ *trajectories.*

during prophase the Cohesin rings are opened by separarases, leaving the place to Condensins: Cohesins are free to move in a wider and less crowded space as DNA is compacted in Mitotic chromosomes (see fig 3.18) (Nasmyth, 2001). The gain in accessible space is most likely the main reason behind the smaller fraction of bound molecules in Mitosis with respect to the control. In panel B of the same figure the reader can appreciate the two Survival Probabilities for Cohesin in Sororin-lacking cells. Already in S/G2 the Survival

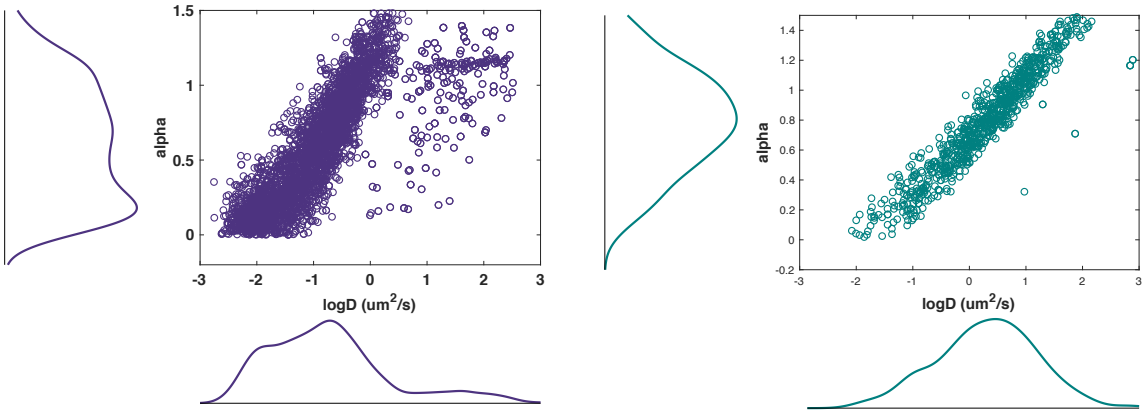



***Figure 3.20: Scatter plot of*** $\alpha$ ***vs*** $logD$ ***for Cohesin in absence of Sororin.*** *Cohesin in absence of Sororin in S/G2 (panel C, violet) and in Mitosis (panel D, emerald). Selected trajectories with* $> 10 frames$*. S/G2 cells:* $N = 17$ *cells ;* $n = 4021$ *trajectories. Mitotic:* $N = 18$ *cells;* $n = 849$ *trajectories. Selected trajectories with* $> 10 frames$*.*

drops more rapidly than in the examples presented before (see fig 3.9 and fig 3.13 for comparison). Sticking to the idea that only the acetylated Cohesins are stabilised by Sororin, such result indicates that *cohesive* Cohesins are actually a subpopulation that we are capturing with our imaging. This may also be a sign of the fact that despite the cell-cycle arrest induced in the Astrocytes, actors like Sororin, that are crucial for sister chromatid cohesion, are still present and accomplishing their task. The Survival Probability distribution for Cohesin in Mitosis confirms the scenario of freely diffusing Cohesins as the association times recorder are of the order of fractions of seconds. The histograms of the logarithm of the apparent diffusion coefficient are shown in panels C and D of fig 3.19. Interestingly, upon Sororin depletion in S/G2 phase Cohesin looses not only more than a half of the stably bound molecules (i.e. $-2 < logD < 1$), but also the freely diffusing population (i.e. $0 < logD < 1$). This controversial result is not of easy interpretation in the context described; a deeper investigation and reflexion are needed. Nonetheless, the histogram of Cohesin in Mitosis strengthen the picture drawn with the other results: the vast majority of molecules display a high diffusion coefficient (i.e. $0 < logD < 1$).

Results will be discussed in section 3.6

### 3.5.2 Cohesin in absence of Nipbl (Scc2)

This last section of results, is dedicated to the investigation of Cohesin's dynamics in absence of Nipbl (Scc2). Following the approach of the previous paragraph, I will first describe Nipbl and its role in the context of the Cohesin complex and then I will illustrate the observations.

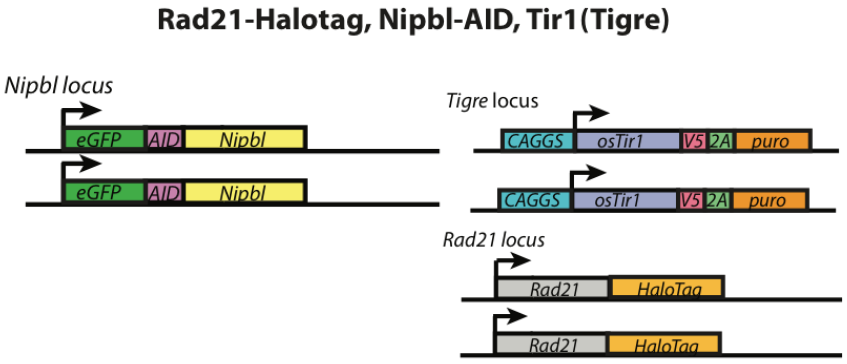



***Figure 3.21:*** *textbfGenotype of the cell line for Cohesin tracking in absence of Nipbl.*

Nipbl is part of the Cohesin loading complex which consist of Scc2 (Nipbl) and Scc4 (Mau2). Nipbl is indispensable to load the complex on chromatin, but it is not needed for sister chromatin cohesion (Ciosk et al., 2000) (Murayama and Uhlmann, 2014). Nipbl is also known to trigger Cohesin's ability to hydrolyse ATP, and it has been recently shown that Nipbl is necessary and sufficient to stimulate Cohesin's ATPase activity in presence of DNA (Petela et al., 2018).
As for the other mutants, Nipbl depletion was achieved via the degron system described in section 2.4.1. Nibpl depletion has more dramatic effects on cell viability

Surprisingly the fraction of bound Cohesins is not significantly affected by the depletion of Nipbl. As Nipbl is required for topological loading of Cohesin on chromatin, this results could indicate that we are actually detecting non-specific Cohesin-DNA interactions, meaning interactions that do not require topological loading and not related to loop extrusion nor sister chromatid cohesion. However, Cohesin's Survival Probability decays drastically in absence of Nibpl, which contradict the idea of pure non-specific interactions. It is possible that the high framerate chosen to quantify the fraction of bound molecules (i.e. $197Hz$) implicate an oversampling of the

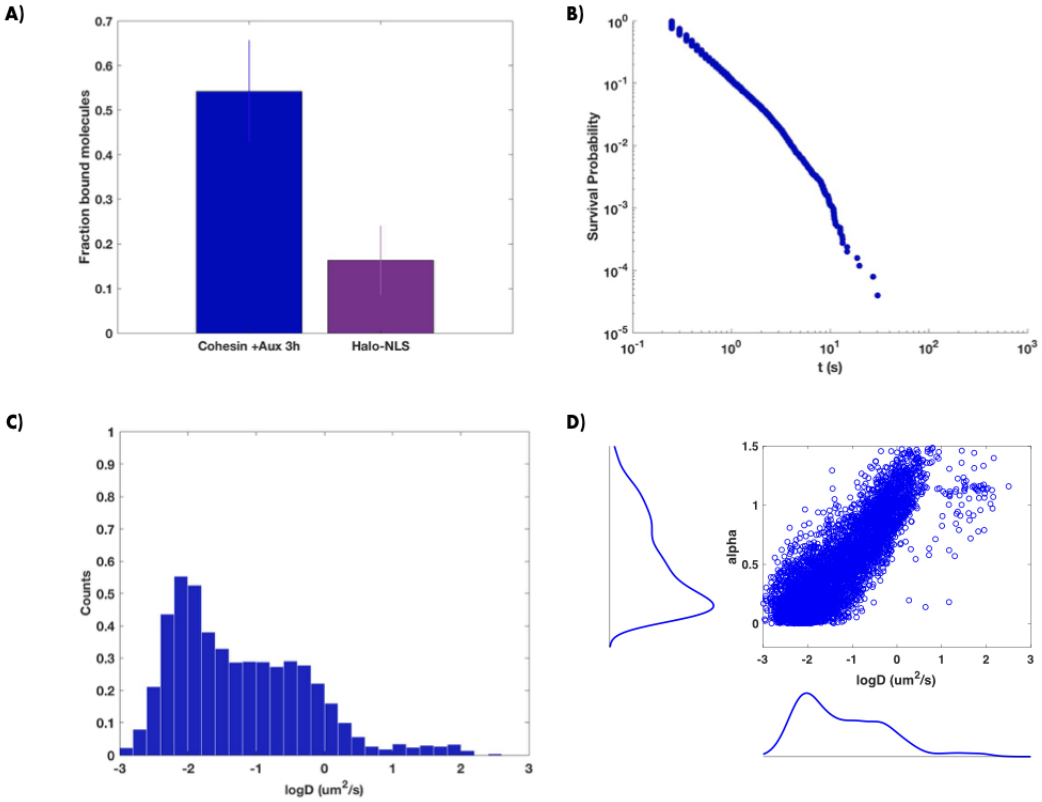



***Figure 3.22: Results of single molecule tracking of Cohesin in absence of Nipbl.*** ***A*** *Fraction of bound molecules obtained from data acquired at* $197Hz$*. N = 23 cells, n = 71696 trajectories.* ***B*** *Survival Probability of Cohesins in absence of Nipbl, from data acquired with* $t_{exp}50ms$*. N = 17 cells, n = 8537 trajectories.* ***C*** *Histogram of the log10 of the apparent diffusion coefficient D, (data with* $t_{exp}50ms$*).* ***D*** *Scatter histogram of the values of the anomalous exponent* $\alpha$ *vs logD. For both C and D, selected trajectories with* $> 10$ *frames. N = 17 cells, n = 6324 trajectories.*

transient interactions.
The distribution of $logD$ is also quite surprising as we observe the loss of a considerable portion of the freely diffusing Cohesins, those with $logD > 0$. This result is frankly of difficult interpretation and more experiments as well as other quantification tool are most likely necessary to better characterise Cohesin's behaviour in absence of Nipbl.

### 3.5.3 Control

Our control experiment consist on tracking the Halotag, with the organic dye JF549, coupled to a Nuclear Localisation Signal (NLS).

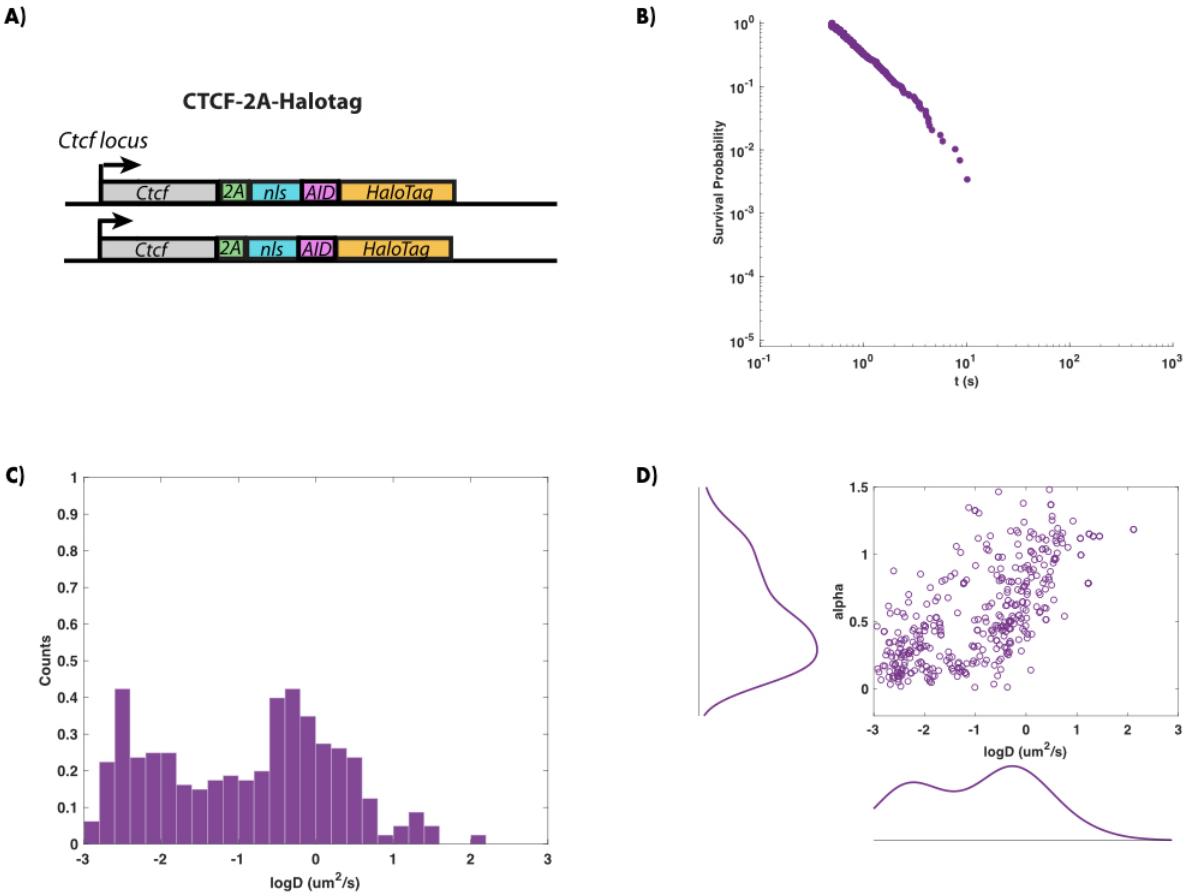



***Figure 3.23: Results of single molecule tracking of Halo-NLS-JF549. A*** *Sketch of the cell line incorporating the 2A cleaving peptide.* ***B*** *Survival Probability of Halo-NLS-JF549, from data acquired with* $t_{exp}50ms$*. N = 5 cells, n = 4758 trajectories.* ***C*** *Histogram of the log10 of the apparent diffusion coefficient D, (data with* $t_{exp}50ms$*).* ***D*** *Scatter histogram of the values of the anomalous exponent* $\alpha$ *vs logD. For both C and D, selected trajectories with* $> 10$ *frames. N = 5 cells, n = 401 trajectories.*

The experiments were performed on the cell line shown in 3.23, panel A. The Halotag-NLS is encoded in the CTCF locus together with a 2A cleaving peptide (Wang et al., 2015). The cleavage happens during translation when,

in our case, the Halotag-NLS is cut from CTCF.
Panel B of fig 3.23 reports the Survival Probability distribution for the control, which decays more rapidly than any other Survival presented in this manuscript, except for the tracking in Mitotic cells. The histogram of the diffusion coefficients shows still a wide distribution that extends to very low values which can be a symptom of transient interactions (panel C). It has been reported by other groups that the Halo-NLS can transiently bind and/or interact (one example is reported by (Hansen et al., 2017). Despite this very short interactions, there is room for a clear distinction between these fast and transient events and those performed by a protein as the Halo-NLS. Tracking of Halo-NLS in other systems (tracking of TALE proteins in U2OS, personal work) showed a very rapid decay at the scale of fractions of seconds. It is possible that in the control used for this work the cleaving system is not 100% efficient and that the few longer interaction, at the scale of the second, are performed by a truncated version of CTCF. In this sense, the tracking of Cohesin in Mitotic cells offers a good control for the identification of the freely diffusion subpopulation. On the other hand the results in Mitotic cells reflect a very peculiar situation where the chromatin, the most crowding element in the nucleus, is highly condensed, leaving more space to the protein to diffuse. As the perspective of Cohesin characterisation will displace the time-scale of interest towards the tens of minutes time-scale I believe the controls presented are reliable.

## 3.6 Discussion

In this chapter I will resume the results presented in the previous sections and discuss them in relation to the state of the art knowledge in the field.

### 3.6.1 CTCF

In the section dedicated to the results obtained for CTCF (sec 3.2), I referred totwo publications of single molecule works on CTCF that appeared in 2017: (Hansen et al., 2017) and (Agarwal et al., 2017). The biological conditions in which CTCF has been investigated in the three different works, the two cited above and the one here presented, are not the same. Hansen and colleagues studied CTCF in mouse Embryonic Stem Cells (mESC) and U2OS cells, both cell lines were edited via CRISPR-Cas9 and stably expressing CTCF coupled to the HaloTag on both alleles. Agarwal and colleagues imaged CTCF in WI-38 cells lines stably expressing CTCF-Halo whose expression

depended on a doxycyclin inducible promoter. The experiments presented in this work were performed in mESC stably expressing CTCF-Halo, with homozygous insertions (for more details see 2.3). While in our work and Hansen's the expression level are endogenous and reasonably homogeneous throughout the population, in Agarwal the expression is triggered by doxycyclin, thus more prone to variations.

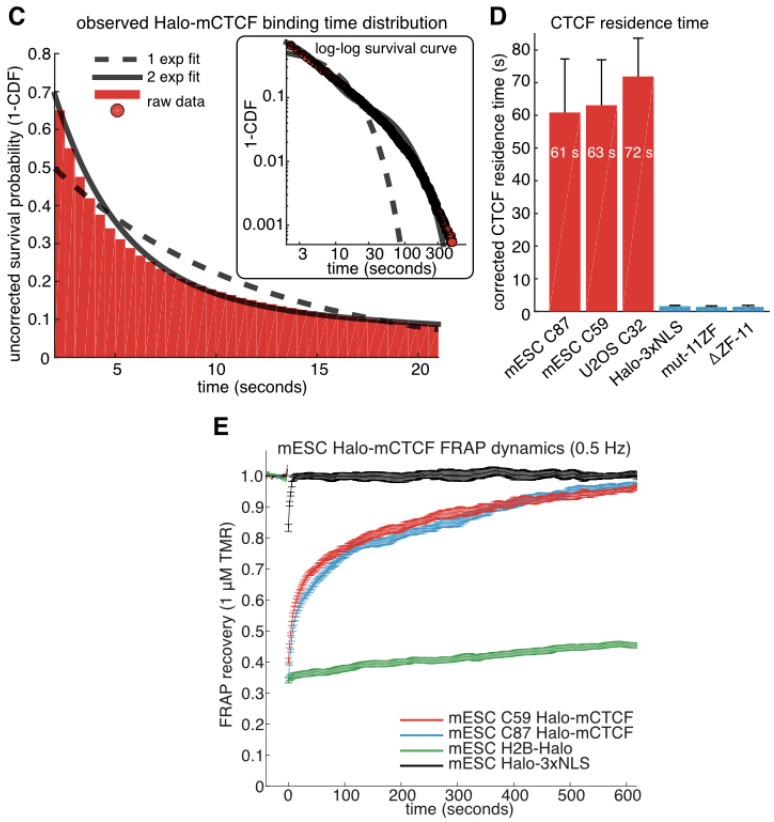
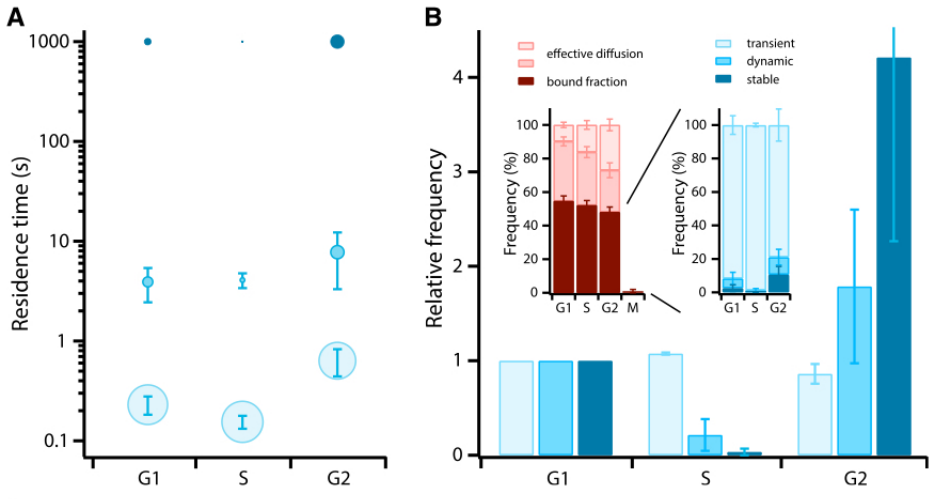



***Figure 3.24: Single molecule tracking of CTCF from (Hansen et al., 2017) .***

Both works report that 50% of CTCF molecules are bound, coherently with what presented in fig 3.2. For what concerns mobility, Hansen reports a value of $\sim 2\mu m^2/s$ for the apparent diffusion coefficient of the freely diffusing population of CTCF. Our results are in good agreement as the peak of the freely diffusing population in the histogram of the logarithm of the apparent diffusion coefficient is $logD \sim 0.25$ which means $D \sim 1.8\mu m^2/s$.
The quantification of residence times follow more or less the same approach in the different works and the differences lay in the interpretation. Neverthe-




C

| | G1 | S | G2 |
|---|---|---|---|
| bound fraction | (54.6 ± 3.1) % | (52.1 ± 2.9) % | (48.1 ± 3.0) % |
| transient binding | (91.7 ± 5.6) %<br>(0.23 ± 0.05) s | (98.7 ± 1.1) %<br>(0.15 ± 0.02) s | (79.0 ± 9.6) %<br>(0.63 ± 0.19) s |
| dynamic binding | (5.8 ± 3.7) %<br>(3.91 ± 1.47) s | (1.2 ± 1.0) %<br>(4.08 ± 0.70) s | (10.4 ± 4.7) %<br>(7.76 ± 4.46) s |
| stable binding | (2.5 ± 1.9) %<br>1000 s | (0.1 ± 0.1) %<br>1000 s | (10.6 ± 4.9) %<br>1000 s |

***Figure 3.25: Single molecule tracking of CTCF from (Agarwal et al., 2017) .***

less the Survival Probability distribution presented in this work (see fig 3.2, panel B) does not reach the values reported in the two other publications because it comes from shorter experiments. In our case, data were acquired with an exposure time $t_{exp} = 50ms$ and movies last up to 4 minutes, the bleaching timescale being the one shown in 2.13. The values reported in (Hansen et al., 2017) are extended to experiments lasting up to 20 minutes and with an exposure of $t_{exp} = 500ms$, while in (Agarwal et al., 2017) Agarwal and colleagues chose to keep the exposure constant at $50ms$ and vary the wait time in their time-lapse experiments. In the two cited works based on CTCF single molecule tracking, the authors identify either 2 (Hansen et al., 2017) or 3 (Agarwal et al., 2017) different dissociation constants. The choice between 2, 3 or more dynamic subpopulations is quite arbitrary but fair as long as motivated. In the roughest approximation we can for sure identify two sharply distinguished populations: bound and freely diffusing. The delicate matter is what to consider bound and how to quantify residence times

and dissociation constants. Hansen choosed to fit his Survival Probability with a double exponential, as shown in fig 3.24 panel C. The residence times shown in panel D are not inferred from the Survival Probability though, but from the histogram of the step length distribution as a function of time (see 2.5.1).

Interestingly the very long binding events, lasting more than 100 s are less than 2%, a very small fraction. On the other hand the FRAP curves shown in panel E of the same figure, report a missing fraction of bound moelcules of $\sim 20\%$ at the same time point of 100 s. This discrepancy is most likely due to the intrinsic limitations of single molecule imaging for very stable events.

Importantly, Agarwal quantified the relative fractions of binding events, reported in Panel A of fig 3.25 and represented by the area of the circles: clearly the so-called transient events (the shortest) dominate the picture. The relative fractions are detailed in the table reported in 3.25 panel C: even if the absolute value of the stably bound fraction is significantly higher than what found by Hansen and colleagues, it still represents a tiny portion of the bound molecules.

Overall, the qualitative picture we get from single molecule tracking of CTCF is coherent amongst the works presented and cited: half of CTCF molecules are bound from G1 to S phase and that the majority of binding events are "rapid" (from 0.1s to 10s of seconds). Differences in the absolute numbers (dissociation constants, number of populations) are probably due to the different biological systems.

None of these studies could directly link CTCF stable binding events to loop extrusion. As CTCF has tens of thousands of binding sites on the genome, and accomplish multiple and different functions, we cannot make conclusion on its dynamics in the context of loop extrusion.

### 3.6.2 Cohesin

The first consideration I will make on Cohesin tracking concerns the characterisation of its dynamics in WT. The slight difference between the tracking performed in the cell line with just Cohesin-Halo (EN130.1) and the untreated cell lines with Cohesin-Halo CTCF-AID (EN131.1) is probably due to a mild biological perturbation caused by the auxin Inducible cassette. Even though in absence of auxin there is no effect on cell viability and chromatin insulation, this cannot rule out small perturbation of the protein stability. The

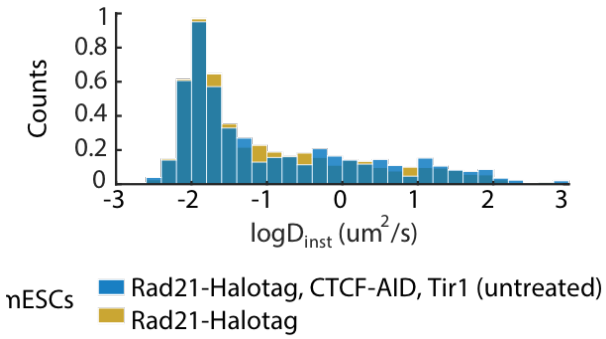



***Figure 3.26: Merge of the distribution of*** $logD$ ***for Cohesin in the Cohesin-Halo and the Cohesin-Halo CTCF-AID (untreated) cell lines.***

plots of $logD$ for the two cell lines mentioned above are merged in fig 3.26.

The various results presented in this work show clearly that CTCF plays no role in the regulation of Cohesin binding kinetics below the minute timescale. This statement is supported by the quantification of the fraction of bound molecules, performed on the data acquired at a frame rate of $\sim 200Hz$ ($197Hz$ precisely) and resumed in fig. 3.28.
The conclusion that Cohesin binding kinetics is not regulated by CTCF has been strikingly validated with ChIP-Seq assay presented in fig 3.27. In absence of CTCF the peaks of Cohesin at the CTCF binding sites are lost, but Cohesin ability to bind chromatin is not affected. The ChIP-Seq has been spike-in calibrated thus we can assert that the amount of Cohesin on chromatin in the cells treated with auxin is the same as the value for untreated cells.

The fact that the fraction of bound Cohesins is not affected in absence of CTCF means that CTCF does not participate in the loading or unloading process of Cohesin, which is not surprising in 2019. As described in section 1.3.2, there are proteins dedicated to this task: in particular Nipbl/Mau2, which are thought to be responsible for Cohesin binding (or loading), and Wapl that triggers the unbinding. Nevertheless, it would be reasonable to think that CTCF plays a stabilising role for Cohesin by interacting with it at its convergent target sites. For this reason the observation that the residence time of Cohesin is not affected by the degradation of CTCF is of less trivial interpretation.

To understand the fact that Cohesins residence time distribution, represented by the Survival Probability, is unchanged upon CTCF depletion it is useful to introduce some considerations related to the loop extrusion hypothesis, as it is in this framework that CTCF and Cohesin interaction is

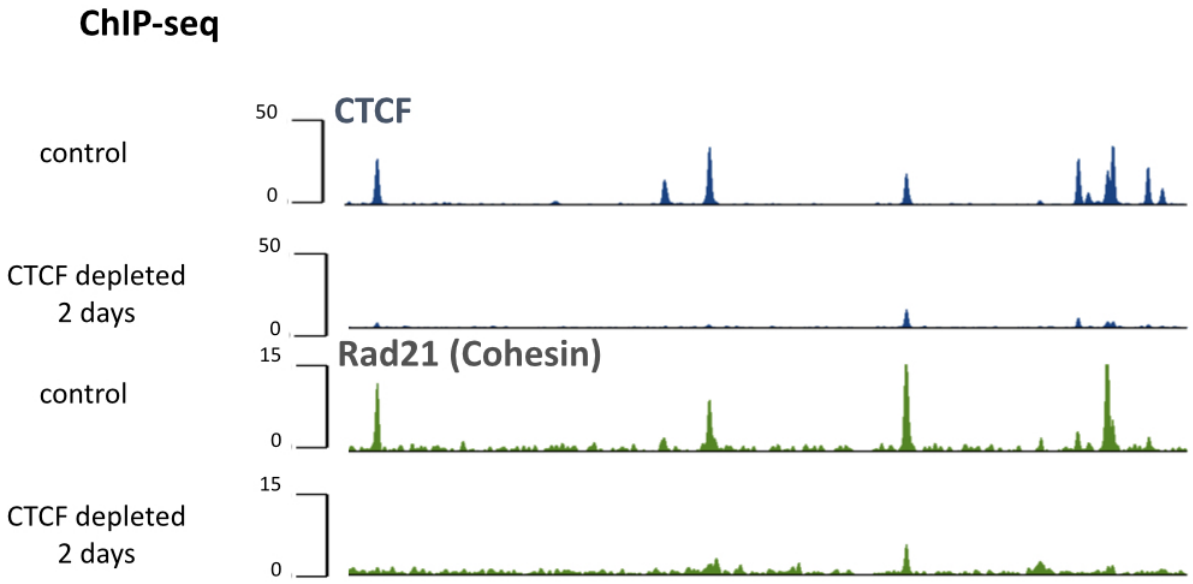



***Figure 3.27: ChIP-Seq of CTCF and Cohesin*** *In wild type cells Cohesin accumulates at CTCF binding sites. When depleting CTCF, Cohesin enrichment is lost but its capability of binding chromatin is not affected. The ChIP-Seq assay is spike-in calibrated. Unpublished data from Elphège P. Nora*

crucial. If we assume Cohesin to have an extruding velocity comparable to the one reported in vitro for an extruding Condensin, i.e. $0.6kbp/s$ (Ganji et al., 2018), and if we consider a rather big TAD of 1.2 Mbp, such as the one imaged in (Bintu et al., 2018) it would take more than 30 minutes to extrude it. If we consider the limit of a small TAD of 100kbp it would still take more than 3 minutes for one single Cohesin to extrude. This calculation supports the hypothesis that we may not be capturing the time-scale required to observe an effect upon CTCF depletion.

Another explanation for the lack of strong difference in Cohesin's behaviour in absence of CTCF could be related to the dynamics of loops.

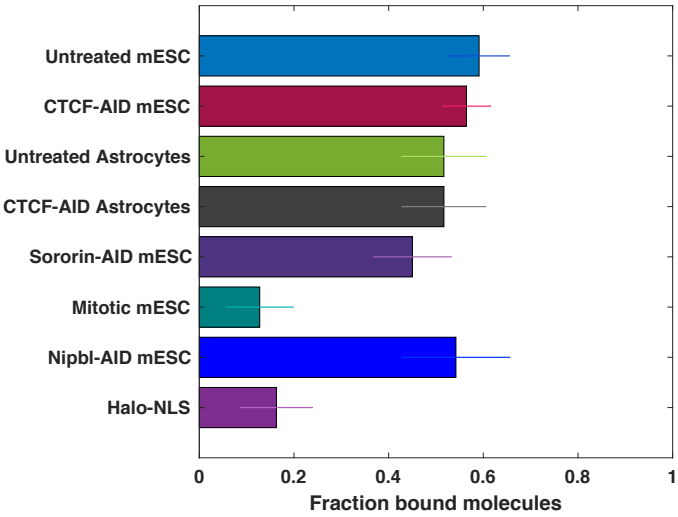
There is still no clear characterisation of the lifetime of a chromatin loop, it is known that they are rapidly lost when Cohesin is degraded (Vian et al., 2018) but these observations arise from Hi-C experiments, which may not be the finest technique to infer dynamic information. Furthermore, the data obtained with single cell Hi-C and super-resolution imaging report a huge cell-to-cell variability, which could be a symptom of a rapid dynamics of loops.

As already mentioned, Single Particle Tracking (SPT), as presented in this work, is strongly limited by bleaching, investigating highly stable events with SPT is very challenging. Time-lapse experiments, stroboscopic imaging, or longer exposures with lower laser powers are in principle a strategy to overcome the photobleaching limitation, yet at longer timescales other issues arise. I performed stroboscopic illumination experiments but could not obtain a clean tracking as the cells start moving after a few minutes of illumination ($\sim$ 3 minutes). The main issue with cell movement is that cells are not just simply drifting (drift is an aberration that could be easily corrected) but the nuclei are considerably deformed. One of the further efforts will be the development of tracking tools that can account for such effect.

Different works have demonstrated the existence of a very stable subpopulation of Cohesins that can stay bound tens of minutes ($>$ 30 min) but there is no trace of such events in the Survival curves. Of course our imaging does not reach such timescales but it would be reasonable to expect an effect of this long tail in the slope of the Survival Probability. A comparison of FRAP and SPT , on the same time-scale, is provided by (Hansen et al., 2017) for CTCF. As already mentioned in the previous section, it is puzzling to observe that the the fractions of stably bound molecules reported by FRAP and SPT have more than a factor 10 of difference. This stands as a warning sign for quantifications. With this *caveat* in mind, another further step will be FRAP experiments in presence and absence of CTCF; by doing so I will be able to validate the published observations in FRAP for Cohesin and study the effects of CTCF depletion.

The mild effect on Cohesin diffusivity in absence of CTCF is conserved amongst Stem cells and Astrocytes. We think that *cohesive* Cohesins (i.e. acetylated Cohesin) in cycling ES cells do not contribute differently to the diffusion parameters that we are measuring than non-*cohesive* Cohesin. Such result could be explained if *cohesive* Cohesin represents only a small fraction of the Cohesins imaged in the S/G2 ES cells, or if the diffusion parameters at the scales we characterised are similar to non-*cohesive* Cohesin.

The interpretation of the results issue from the tracking of Cohesin in the



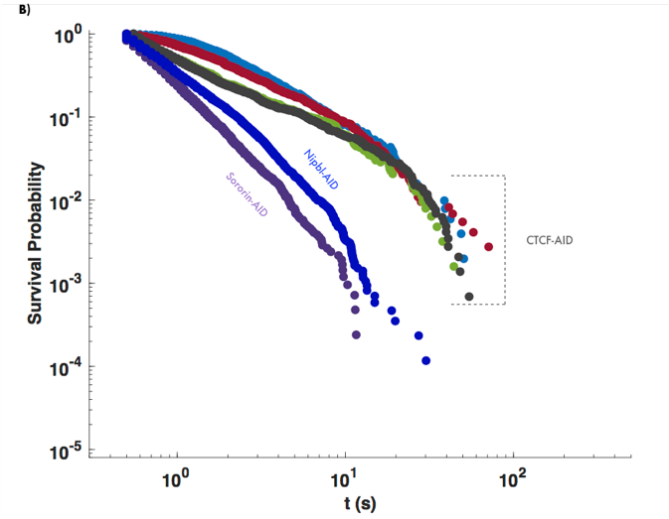




***Figure 3.28: Recapitulated results for the fraction of bound Cohesins (left) and its Survival Probability (right) in the context of different alterations.***

context of other alterations, in particular in absence of Sororin or Nipbl, is more intricate. It is important to mention that, differently from the other cell lines, the experiments presented for Cohesin-Halo Sororin-AID and Cohesin-Halo Nipbl-AID represent only one replicate. The discussion that follows is then preliminary and needs to be confirmed by larger statistics.

I will start by discussing the observations reported in absence of Sororin.

When depleting Sororin the fraction of bound molecules is slightly decreased to$\sim$ 45% for cells in S/G2. Since Sororin is involved in the stabilisation of uniquely the acetylated Cohesins, it is reasonable to think that the small loss of bound molecules represent the loss of stabilised acetylated Cohesins. This results is in accordance with what observed in non-cycling Astrocytes, where we do not expect Cohesins to be stabilised and we observe a lower fraction of bound molecules (50%) than what measured in S/G2 (60%). The difference between ES and Astrocytes suggests that the acetylated Cohesins in our system are $\sim$ 10%, which is the loss observed when degrading Sororin. Cohesin's residence time distribution (represented by the Survival Probability) is also affected by Sororin depletion, as shown in fig 3.28, The dynamics of Cohesin's acetylation and stabilisation by Sororin is not yet completely understood thus we cannot make any strong statement in this regard.

Upon depletion of Nipbl we observe no significant effect on the fraction of bound molecules. This unexpected result may be explained by the fact that the majority of interactions sampled at $\sim 200Hz$ are transient interactions, even though this argument would weaken the conclusions made for the alterations previously described. Interestingly the Survival Probability is affected by the degradation of Nipbl: this residence time distribution decays definitely faster than the case of Cohesin +/- CTCF but slower than the Sororin-AID mutant. This result may be representing the loss of the topologically loaded Cohesins, but needs to be validates with more statistics as well as other experimental approaches such as FRAP. Furthermore, this could indicate that the loading activity of Nipbl is involved in stabilizing Cohesin on chromatin, so that in the context of Nipbl degradation we observe more Cohesin molecules that are not topologically loaded, and therefore have a faster diffusion rate.The most puzzling results is provided by the histogram of the apparent diffusion coefficient. Contrary to any prevision when depleting Nipbl, Cohesin's loader, we observe a considerable decrease of the mobile fraction. Before speculating on possible mechanisms behind this result, more and diverse experiments are needed.

## 3.7 Conclusion and perspectives

In this manuscript I presented a characterisation of CTCF and Cohesin dynamics in living mouse embryonic stem cells. The individual behaviour of the two proteins and their interplay were studied with a single molecule track-

ing approach. The main take home message of this work, concerning the interplay between CTCF and Cohesin, is that CTCF plays a crucial role for Cohesin positioning on chromatin, but has no role on its stability up to the minute time-scale. Cohesin's dynamics is mildly affected by Cohesin depletion but finer analysis tools are needed to better characterise such effect.
To push further this investigation it is crucial to image Cohesin on longer time-scale, both with single molecule and bulk approaches (FRAP).
Another aspect I would investigate is the chromatin organisation in the absence of CTCF. In particular I would track the most stable cohesins in a bulky regime, use them as a marker of plenty of genomic loci and observe weather CTCF depletion induces changings in chromatin mobility or arrangement, following the approach described in (Zidovska et al., 2013).
Last but not least I would also pursue a set of Stochastic Optical Reconstruction Microscopy (STORM) experiments that I started, to study if and how CTCF depletion affects Cohesin positioning in the entire nucleus.
The results found for Cohesin in absence of Sororin and in absence of Nipbl will be tested with new acquisition and larger statistics, in the near future. The single particle tracking experiments for Nipbl will be performed as a function of incubation time to see how the remaining amount of Nipbl may affect Cohesin's behaviour. Furthermore FRAP experiments will be performed for the Nipbl-AID cell line to explore longer timescales.
The experimental findings will be corroborated with further simulations in collaboration with the group of Leonid Mirny at MIT, Cambridge, MA. Simon Grosse-Holz is currently exploring different physical models to interpret our characterisation of Cohesin's mobility and residence time. In particular we seek eventual signatures of active motion on chromatin.

Overall, the results of this work cannot be seen as a proof or a disclaimer of the loop extrusion hypothesis. The ChIP-Seq assay points definitely to the role of CTCF in the loop positioning. On the other hand, the mild effects on Cohesin's behaviour in absence of CTCF seem to point towards the opposite direction as, according to loop extrusion, we would expect a stronger impact upon CTCF depletion. Various aspects of the loop extrusion model are not set yet (for example we still don't know weather Cohesin slides on chromatin while extruding or stays almost immobile while the fibre passes through the ring) and the dynamics of loops is still poorly understood. I believe we should not think of this model as an untouchable ground truth but as a solid base to investigate and understand chromatin spatial organisation. The next years will be extremely exciting in this sense. I am persuaded that the combination of the efforts in polymer modelling, and the rapid development of cutting edge techniques to resolve chromatin architecture (mainly conformation capture

and super-resolution imaging methods) will significantly shift the paradigm of knowledge in the field.

# Acknowledgements

PhD is not a lonely work. Let me here thank all the people that contributed scientifically, humanly and in whatever manner.
I First want to reaffirm that this work is dedicated to the memory of Maxime Dahan. He was a special man, amazing and unbearable as a PhD supervisor can be. I missed him for this entire year and I realised that the void he left will not be filled, as Maxime was one of a kind (there is only one PhD supervisor in life, like mom). I am deeply proud to be one of his students and I thank him once more for believing in a crazy intern that barely knew how to switch on a microscope. We can't fill the void, but we can point at it.
I would also like to thank all the jury members who accepted to review my work. A big thank goes to Mathieu Coppey, who adopted me as his student for this last run and assured to the entire team guidance and support; his enthusiasm and scientific creativity motivated me to do better and keep going.
Thanks to the LOCCO team (Light Observation and Control of Cellular Organisation (that's the first time I manage to write it without checking on our webpage.)), a special thanks to Elie, a rare good friend. I owe four years of gratitude to Bassam Hajj, a.k.a. *Chef*. We have been office mates for four years and we faced together quite a few big life turning points. I will avoid reporting crazy anecdotes to save him the embarrassment, but let me affirm how a great scientist and person he is. A mentor, a great friend: family. Chef, you won't get rid of me, surrender.
It is hard to find the words to thank Elphège Nora, nothing of what I have been doing in the last years could have been possible without him. But he went further, from the beginning Elphège showed me how to be an excellent scientist: brilliant, honest, meticulous, critical but always constructive and incredibly wise. I feel extremely lucky for being the single molecule girl when he knocked on Maxime's door a few years ago, with these amazing cell lines. Beyond science, Elphège was a pillar during this hard year without Maxime. He was always there and supported me unconditionally. I will never find the

words to thank him enough.
I want to also thank Elphège post-doc mentor, Benoit Bruneau, who supported this project from the beginning and made it possible; I wish to thank him also for the great support he showed me in the last year.
*Merci beaucoup* to Antoine Coulon, with whom I shared the first steps of this project and who showed me, from the beginning, how to do good science. Thanks Antoine for the support and for pushing me to do better. I wish you the best and hope our path will stay crossed.
Leonid Mirny is like a putative supervisor to me, and more. I would like to thank him for a thousand little moments of exchange and support that went, luckily, way beyond science. Thanks to Johannes Nuebler for the great discussions and for his contagious passion for science
A big thank you goes to Angela Taddei, a model and an inspiration for young women in science. Thanks for all the discussions and the support. I also thank Raphael Voituriez who, together with Angela, accepted to be part of my thesis committee and always has great scientific insights that come with a smile.
Thanks a lot to Agnès LeSaux and Julia Ronsch, from the team of Edith Heard, who patiently trained me to deal with stem cells and helped me on a list of other biological issues during these years. And thanks to Edith Heard who not only shared her cell lines, knowledge and lab engineers, but who also let me feel support and guidance in this last year. Thanks to Luke Lavis for generously providing the fundamentals of our daily routine: the JF dyes.
I would also like to thank Maud Bongaerts from our team, who quickly learned to handle stem cells and helped me with the cell management in the last year.Thanks to Anders Hansen for the insightful discussions and his kindness.Thanks to Simon Grosse-Holz, for his enthusiasm and his wonderful simulations.
Thanks to Alexandra Elbakyan and her collaborators for the idea and the realisation of Sci-hub, and for their brave resistance. I should not mention how many papers I had access to thanks to their great work. Science belongs to the curious.
A big thanks to my family: thanks to my dad, who left us too early, this is also for you. Thanks to my mother who taught me strength and resilience, thanks to my sister who is just incredible. Thanks to my beloved Louis, for motivating me to live a happy life. I lost quite some important persons on this path, but on the other hand I found a bunch of wonderful friends, colleagues and mentors. And I feel extremely lucky to have met so many great scientist, so many wonderful human beings. I am very grateful for this amazing experience and I feel this is just the beginning. Thanks to you all guys. Science is a choral effort, and that is what makes it fun!

## 3.8 Résumé en français

Résumé du manuscrit en frans.

Les pages suivantes résument l'introduction et les résultats obtenu pendant ma thèse et sont présenté en frans comme demandé pas l'école doctorale de physique Ile-de-France. C'est curieux qu'une uvre dédiée á un public bien habitué á s'exprimer en anglais, la communauté scientifique, doit e traduit. J'espère que cet exercice soit remis en question dans le futur.

Le noyau est un environnement très complexe et encombré. C'est le centre o les processus cellulaires importants sont réglementés. Cet environnement encombré est cependant très organisé. La complexité de ce système en faisait un cauchemar de biologiste et un terrain de jeu de physiciens. Au cours de ma thèse, j'ai abordé diverses questions difficiles liées aux mécanismes qui se produisent dans le noyau, principalement avec des techniques d'imagerie á une seule molécule. Ci-après, je présente quelques-uns des projets dans lesquels j'ai été impliquée et développe davantage le sujet principal de mon travail décrit dans cette thèse. J'ai d'abord abordé le noyau des mammifères au cours de mon stage, lorsque j'ai commencé á étudier le processus de recherche de cibles de protéines de liaison á l'ADN, telles que les facteurs de transcription. Lors de la recherche de leur cible, les facteurs de transcription interagissent de manière non spécifique avec diverses séquences d'ADN. Pour mieux comprendre ce processus, j'ai effectué des expériences de suivi d'une molécule unique sur des protéines ressemblant á des facteurs de transcription. Pour imiter les facteurs de transcription, j'ai travaillé avec les protéines TALE (Transcription Activator Like Effectors), qui peuvent e cons pour cibler toute séquence d'intérn modifiant deux résidus centraux (les 12 et 13) dans le domaine de liaison á l'ADN. Gr á la polyvalence de cette méthode, 6 séquences TALE ciblant des séquences de différentes longueurs ont été cons et étudiées. En particulier, nous avons caractérisé les protéines avec des séquences apparentées de 6 pb (paires de bases), 10 pb, 14 pb, 18 pb, 22 pb et 26 pb. Les interactions non spécifiques, révélées par l'imagerie et le suivi d'une molécule unique, ne sont pas affectées de manière spectaculaire par la longueur du domaine de liaison á l'ADN de la protéine, á l'exception du plus court (6 pb et 10 pb) qui diffère des autres échantillons. Les répartitions des temps de résidence sont toujours caractérisées par une longue queue (loi de puissance) indiquant que la stabilité de l'interaction non spécifique ne dépend pas de la longueur du domaine de liaison á l'ADN. Notre interprétation de cette observation est que, bien que les protéines courtes subissent des interactions non spécifiques en raison de la redondance élevée des séquences courtes dans le génome, les protéines plus longues sont plus tolérantes aux défauts d'appariement et peuvent donc e kidnappées par des séquences quasi-apparentées quelques nucléotides sont

suffisants sur le plan énergétique pour effectuer une liaison transitoire). Ce sujet ne sera pas traité dans ce manuscrit, car il nécessiterait une discussion longue et détaillée. Nous terminons actuellement ce travail et espérons que nos résultats seront bientt publiés. L' étude de la recherche de cibles dans le noyau de cellules de mammifères pose la question de l'accessibilité de la chromatine et c'est ainsi que je me suis intéressé á l'organisation spatiale et á la dynamique de la chromatine.

Poussée par la curiosité pour l'organisation nucléaire, j'ai commencé á rechercher des loci génétiques, principalement dans des cellules souches embryonnaires de souris. Nous avons entamé une collaboration enrichissante avec le professeur Edith Heard et son équipe, ce qui m'a amené á étudier le comportement dynamique des locus á des distances génomiques connues et á différents états de transcription (actif / silencieux) sur le chromosome X. Les résultats préliminaires sont prometteurs mais, comme ce projet est toujours en cours, une analyse plus approfondie est encore nécessaire avant de pouvoir tirer des conclusions.

Ce manuscrit récapitule les résultats du projet principal mené durant ma thèse: une étude molécule unique de l'interaction entre deux facteurs nucléaires qui jouent un rle crucial dans la régulation de l'architecture de la chromatine, CTCF (facteur de liaison au CCCTC) et Cohesin. Ce sujet sera traité en détail dans les chapitres de ce manuscrit. La thèse est divisée en 4 chapitres; je décris désormais le contenu de chacun.

Le chapitre 1 est consacré á l'introduction du lien étroit entre structure et fonction. Après une description des avantages découlant du repliement tridimensionnel et dynamique du génome, je décris les principales techniques utilisées pour capturer la structure de la chromatine. Un bref aperst ensuite présenté sur l'organisation multi-échelle de l'ADN en mettant l'accent sur les domaines d'association topologique (TAD), qui constituent l'échelle d'intére cette étude. Le paragraphe suivant est consacré á la description d'une hypothèse de mécanisme de formation de TAD de plus en plus crédible: l'hypothèse de l'extrusion en boucle. Le chapitre se termine ensuite par deux paragraphes sur les protagonistes du modèle d'extrusion de boucle et l'objet de cette étude, le facteur de liaison CCTC (CTCF) et Cohesin. Le chapitre 2 décrit les techniques déployées dans le but de ce travail. Une brève introduction á la microscopie optique est présentée, suivie d'un paragraphe sur la microscopie de localisation d'une molécule. Une description détaillée des lignées cellulaires et des conditions de culture cellulaire est également fournie. La dernière section de ce chapitre est consacrée á une description détaillée de l'analyse des données de suivi d'une molécule unique. Le chapitre 3 présente les résultats du suivi par molécule unique des différentes protéines, dans le contexte de diverses modifications. La première section est un bref aperu

suivi d'un locus génomique utilisé comme référence pour la dynamique de la matrice (chromatine). Cette section est suivie des résultats du suivi du CTCF et des résultats pour la Cohesin en présence ou en l'absence du CTCF, en l'absence de Sororin et en l'absence de Nipbl (deux facteurs impliqués dans la régulation de la cinétique de liaison á la Cohesin qui seront présentés en détail). dans leur section respective). Le manuscrit se termine par une section dans laquelle tous les résultats sont discutés et placés dans le contexte de la littérature actuelle et des connaissances de pointe en la matière.

Chapitre 1

Nous savons aujourd'hui qu'il existe plusieurs niveaux d'organisation de l'ADN. Le plus petit domaine d'organisation structurelle auquel on puisse penser (chez les eucaryotes) est le nucléosome: environ 150 pb d'ADN enveloppé / sont enveloppés d'histones. á l'échelle des dizaines á des centaines de kbp, le génome est organisé en domaines appelés topologiquement associer domaines (TAD), domaines caractérisés par des interactions intra-domaines plus élevées plutt qu'entre domaines, pour les génomes de mammifères. Sur l'échelle génomique plus large de Mbp, nous trouvons des compartiments qui correspondent aux régions euchromatiques et hétérochromatiques citées dans la section 1.1. Ces domaines peuvent e soit des compartiments A (riches en gènes, actifs sur le plan transcriptionnel, moins compacts, typiquement situés au centre du noyau), soit des compartiments B (pauvres en gènes, plus compacts, situés á la périphérie du noyau). Le plus grand domaine identifié est le territoire chromosomique; Les chromosomes ne se mnt pas, chacun d'entre eux consiste donc á former son propre domaine.

Les TAD et les compartiments donnent lieu á des motifs spécifiques dans une carte Hi-C. De nombreuses découvertes dans les mécanismes á la base de la régulation de la structure de la chromatine sont liées aux différents phénotypes apparaissant sur ces cartes. Les caractéristiques et les différences des modèles de TAD et de compartiments dans les cartes Hi-C permettent de distinguer les perturbations á chaque niveau d'organisation de la chromatine. Au cours des 2-3 dernières années, nous avons appris par exemple que les facteurs d'appauvrissement impliqués dans la formation de TAD n'ont pas nécessairement d'impact sur les compartiments (Schwarzer et al., 2017), (Nora et al., 2017), (Haarhuis et al., 2017) et (Rao et al., 2017) (Pour une discussion plus détaillée, voir section 1.2). Je traiterai des effets de la modification de la réglementation et de la maintenance des TAD dans la section suivante. Je tiens ici á souligner qu'il est prouvé que les TAD et les compartiments ne résultent pas du m processus physique. Un modèle a récemment été proposé pour expliquer l'interaction des différents mécanismes derrière les TAD et les compartiments (Nuebler et al., 2018). Une quantité informative pour la description de l'arrangement spatial de la chromatine est la proba-

bilité de contact P (s), qui représente la fréquence de contact renormalisée en fonction de la distance génomique s. Il est intéressant de noter que la distribution de probabilité de contact décrouivant une loi de puissance, qui est une fonction intrinsèquement sans échelle. Quelques exemples sont rapportés á la figure 1.10. Par souci de brièveté et de manière cohérente avec les objets de cette étude, je me concentrerai sur les TADs. Dans les sections suivantes, je présenterai d'abord les TADs et les mécanismes proposés pour les réglementer, puis je continuerai en décrivant les acteurs impliqués dans ce processus.

Sur les TADs

J'ai ici l'intention de brosser un tableau de base des caractéristiques des TADs dans les termes les plus généraux; Cette brève description ne prétend pas e exhaustive, mais fournir au lecteur les éléments fondateurs de ce travail. Les TAD sont une découverte récente qui a profondément influencé les domaines connexes. Les trois études phares sur les maladies animales transfrontières ont été publiées en 2012 (Dixon et al., 2012), (Nora et al., 2012), (Sexton et al., 2012). Dans la figure 1.11, le lecteur trouvera une chronologie simplifiée qui marque les jalons de la découverte et de l'enqu TAD. Les domaines d'association topologique (TAD) sont des régions de chromatine auto-interagissant, ou des régions qui ont tendance á interagir davantage avec elles-ms qu'avec d'autres régions. Les TAD représentent une échelle préférentielle de contacts cis- fonctionnels (Zhan et al., 2017) et la perturbation de ce niveau d'organisation peut avoir des conséquences dramatiques.

L'image actuelle des domaines repose sur des tests statiques, mais le TAD ne doit pas e considéré comme une structure stable ou omniprésente. Des expériences sur des cellules uniques ont révélé la forte variabilité entre cellules qui est lissée dans les dosages canoniques de Hi-C (moyenne calculée sur des dizaines de milliers de cellules) (Nagano et al., 2013) (Flyamer et al., 2017).

Le paradigme du repliement de l'ADN a considérablement changé au cours des 10 dernières années. Le problème de la description et de la prévision de l'arrangement de la chromatine dans l'espace a été abordé sous différentes approches. Les chercheurs ont développé des modèles basés sur la physique des polymères purs (Marko et Siggia, 1997) (Rosa et Everaers, 2008), d'autres s'appuyant sur des informations épigénétiques pour définir une interaction préférentielle (Jost et al., 2014) ou sur des informations structurelles (Giorgetti et al. , 2014), ou d'autres construits sur super enroulement d'ADN (Benedetti et al., 2014). Pour ce qui concerne les TAD, un mécanisme proposé attire de plus en plus d'attention et de preuves á son support: l'hypothèse d'extrusion de boucle. L'idée de base est que des protéines spécifiques sont capables d'embrasser la cha d'ADN et de former des boucles; á mesure que les boucles grossissent, des éléments de régulation distaux sont attirés et un

domaine est formé, see Fig 1.14. La figure 1.11 montre les principales étapes qui ont contribué á l'étude de l'interaction entre le CTCF et Cohesin ; ces deux facteurs étaient connus pour jouer un rle "structurel" avant le début des années 2000, mais aucune discussion contradictoire n'a été envisagée. Au cours de la dernière décennie, différents groupes ont démontré que le CTCF et Cohesin sont impliqués dans les ms processus (Kagey et al., 2010).

Le CTCF, le facteur de liaison du CCCTC, est un doigt de 11-zinc, un protéine liant l'ADN avec des dizaines de milliers de sites cibles spécifiques sur le génome (Ohlsson et al., 2001). La FCCC a été classiquement décrite comme un isolant, un facteur qui relie les éléments régulateurs distaux en bouclant l'ADN (Phillips et Corces, 2009) 5. On sait aussi que la CTCF est reconnue pour établir un pont entre les interactions entre des domaines génomiques spécifiques en marquant les limites des domaines associés á Lamina (Guelen et al., 2008), m si la diminution de la CTCF n'affecte pas ces domaines (Nora et al, résultats non publiés). Entre 15 000 et 40 000 sites de liaison ont été identifiés pour la CTCF, tant aux limites qu'á l'intérieur des DAT, et la séquence cible de la CTCF s'est révélée très conservée pendant l'élution (Kim et al., 2007). Il a été récemment démontré que, chez les mammifères, les limites des DAT sont définies par les sites convergents du CTCF (Rao et al., 2014) (de Wit et al., 2015) (voir aussi figure 1.19). La suppression ou l'inversion de ces sites peut avoir des conséquences dramatiques : déjá dans l'un des travaux d'établissement des DAT, il est démontré que l'altération des sites CTCF implique une perturbation des DAT (Nora et al., 2012). De plus, Lupian?ez et ses collègues ont démontré que l'altération de certains sites de liaison de la CTCF entra un mauvais pliage des DAT et une pathogenèse consécutive (Lupi?an?ez et al., 2015). Le remodelage de domaines comme les TADs implique souvent d'affecter l'expression des gènes, car les perturbations de l'isolation médiées par la CTCF donnent lieu á des interactions enhancer-promoteur différentes des conditions du type sauvage. La corrélation entre la perturbation de la TAD et la maladie a été observée dans différents contextes (Valton et Dekker, 2016), la modification/suppression de séquence CTCF est généralement une condition suffisante pour affecter le compactage de la chromatine á l'échelle TAD. Dans les sections suivantes, et dans le reste du manuscrit, je me concentrerai sur les mammifères, et sur la souris en particulier.

La cohésine est un complexe multiprotéique. Dans sa description minimale, l'anneau de Cohesin se compose de deux protéines appartenant á la famille des Complexes de Maintien de Structure (Smc), Smc1a et Smc3, et d'une sous-unité kleisine, Scc1 également connue sous le nom de Rad21. Smc1a et Smc3 forment un hétérodimère en forme de V fusionné á ce que l'on appelle la charnière (voir Fig 1.16) ; les deux bras libres, avec leur ATPase

sont pontées par Scc1. Les sous-unités et cofacteurs de Cohesin sont énumérés dans le tableau 1.15. Le complexe Cohesin est un grand objet mesurant plus de $\sim 50nm$ de hauteur et jusqu'á $\sim 40 - 50nm$ de largeur (Anderson et al., 2002), comme le montre la microscopie électronique dans la figure 1.16 panneau de droite. Il n'y a pas de structure cristalline complète et définitive pour Cohesin, mais le panneau de gauche de la figure 1.16 représente un scénario probable et donne une idée de la taille de Cohesin par rapport á un nu-cléosome (entre les deux bras Smc). Dans cette image, le lecteur peut également apprécier la présence d'autres facteurs que ceux énumérés ci-dessus : Pds5, Scc3, Wapl, Scc2/Scc4 sont des protéines nécessaires au bon fonctionnement de la Cohésine. Pds5 et Scc3 (SA1 et SA2 chez les mammifères) sont deux supplémentaires du complexe de la Cohésine, les protéines dites HEAT6 répétées associées aux Kleisins (HAWKs), ou protéines HEAT associées aux kleisines (Wells et al., 2017) (il a été démontré que la SA fait partie du complexe Cohesin avec stabilité (Gerlich et al., 2006)). Désormais, en mentionnant Cohesin, je me référerai á l'hétérodimère Smc1/Smc3 avec la kleisine Rad21 et SA1/SA2. Une autre protéine interagissant avec les keisins du complexe, mais de manière transitoire, est le Scc2 (Nipbl). Il a été démontré que le Nipbl est nécessaire pour la charge de co-résine sur la chromatine (Petela et al., 2018) il a également été suggéré qu'une fois Cohesin chargé, le Nipbl saute d'un complexe de Cohesin á l'autre pour, très probablement, stimuler son activité ATPase (Rhodes et al.). Wapl est connu sous le nom de déchargeur de Cohesin, l'une des preuves majeures de cette affirmation étant que la déplétion de Wapl entra un phénotype de vermicelle (chromatine hautement compactée) pour les chromosomes en interphase (Tedeschi et al., 2013) (Wutz et al., 2017)7. Le scénario dans lequel Nipbl stimule l'activité de Cohesin AT- Pase et est en compétition avec Psd5 pour lier le sous-domaine Scc1, est plausible mais pas encore définitif. Un travail particulier a montré que Pds5 et Wapl peuvent favoriser le chargement de Cohesin sur l'ADN en l'absence de Nipbl (Murayama et Uhlmann, 2015) (scénario pour le moins contradictoire compte tenu des nombreuses preuves des effets de l'appauvrissement de Wapl (Tedeschi et al, 2013) (Wutz et al, 2017)). Néanmoins, ce résultat discuté est cohérent avec le fait que l'épuisement des Pds5 entra une perturbation des TAD (Wutz et al., 2017). Jusqu'á il y a quelques années, Cohesin était principalement étudié pour son rle dans la cohésion des chromatides soeurs. Cohesin est l'anneau qui maintient les deux chromatides soeurs ensemble pendant la réplication. Dans G1 Cohesin peut effectuer une liaison transitoire, tandis que dans la phase S une sous-population de Cohesins est acétylée par l'acétyle-transférase Esco1 ou Esco2. L'acétylation verrouille la Cohésine sur la chromatine, ce qui permet d'assurer une bonne cohésion lors de la réplication chromosomique ; la

co-hésion est finalisée lorsque la Sororine a lié le complexe. La sororine est une protéine en compétition pour le site de liaison Wapl sur la Cohésine, son association au complexe est considérée comme ayant un effet stabilisant. Dans Prophase, la sororine se détache et la majeure partie de la cohésine est libérée de la chromatine par la phosphorylation de la sous-unité SA. Pendant ce temps, Shugoshin (SGO1) et PP2A s'accumulent aux centromères pour prévenir localement la dissociation des Cohesins. Enfin, dans l'anaphase, les anneaux Cohesin restants sont ouverts par une Séparase qui coupe Rad21. Le processus est résumé dans la figure 1.17. Pour plus de détails, voici quelques publicités sur le sujet : (Peters et al., 2008) (Nasmyth et Haering, 2009) Losada (2014) (Kanke et al., 2016). Il n'y a pas encore de consensus sur le rle spécifique de chaque facteur impliqué dans le métabolisme de la Cohésine. Des preuves récentes suggèrent que le tableau n'est pas aussi binaire qu'on le pense, que les protéines ne remplissent pas nécessairement une fonction simple (chargement, déchargement, translocation, blocage) (résultats non publiés Nora). Deux travaux très récents ont permis de mieux comprendre le remodelage structurel de Cohesin en association avec l'ADN (Chapard et al., 2019) (Marko et al., 2018). Il en ressort que Cohesin a deux compartiments, l'un entre les bras Smc et l'autre situé entre la sous-unité kleisine (Rad21) et les ts ATPase des protéines Smc. L'ADN est piégé dans ce dernier compartiment comme le montrent les deux travaux résumés á la figure 1.18. Alors que dans (Chapard et al., 2019) le problème est étudié avec une approche biochimique expérimentale (panneau de gauche dans la figure 1.18), la deuxième référence (Marko et al., 2018) est le résultat d'un modèle théorique (panneau de droite dans la figure 1.18). La cohésine a également été partiellement caractérisée par des essais in vitro sur une seule molécule (Stigler et al., 2016) (Davidson et al., 2016) (Kanke et al., 2016). Les expériences présentées dans les trois références sont basées sur la m technique consistant á attacher de nombreux brins d'ADN sur une lamelle, comme un rideau, á mettre la protéine d'intérn solution et éventuellement á appliquer un flux. Les brins d'ADN et la protéine sont détectés en fluorescence, marqués avec des fluorophores de couleurs différentes, ce qui permet de caractériser le comportement du pro-téine, principalement le temps de séjour et le glissement, á l'échelle de la molécule unique. Il a été démontré que Cohesin seul effectue de nombreuses interactions transitoires, que Cohesin glissant sur l'ADN est dépendant de l'ATP. Stigler et ses collègues ont signalé que la présence du facteur de charge de Cohesin Mis4 (l'homologue S.pombe du MAU2 humain) a entrae une augmentation significative du nombre d'événements de liaison, tandis que Davidson et al ont montré que le CTCF limite les translocations de Cohesin. Les deux travaux ont montré que Cohesin est capable de franchir de nombreux obstacles différents (dCas9, EcoRI, Nucleosomes, et autres).

Ces beaux résultats ont été obtenus lors d'un test in vitro, la plupart du temps sur de l'ADN nu et dans différentes conditions salines donc, malgré leur valeur incroyable, ils ne reflètent pas nécessairement ce qui se passe dans une cellule vivante. Plus important encore, nous ne savons pas si Cohesin utilise l'extrusion en boucle pour se déplacer dans ces essais, ou s'il diffuse simplement passivement en suivant le flux tampon exercé dans la configuration expérimentale. Par souci d'exhaustivité, je dois mentionner que Cohesin est également impliqué dans la réparation de l'ADN pendant les phases S et G2 (Wendt et Peters, 2009) mais je ne traiterai pas le sujet puisque celui-ci fait l'objet d'un champ dédié et complexe.

Comme mentionné á la fin de la section précédente, Cohesin est l'anneau qui maintient les deux chromatides soeurs ensemble après la réplication de l'ADN et pendant des décennies il a été étudié en relation avec la réplication de l'ADN et la division cellulaire (Nas- myth et Haering, 2009). Pourtant, ses niveaux d'expression sont considérables m dans les cellules non cycliques, c'est l'une des raisons qui ont incité certains scientifiques á explorer d'autres fonctions possibles du complexe (Wendt et al., 2008). Bientt, on a découvert que Cohesin jouait un rle dans la régulation des gènes en coopération avec la CTCF et avec le complexe Mediator chez les mammifères (Kagey et al., 2010) et chez la drosophile (Pauli et al., 2010). La cohésine a été désignée comme extrudeuse d'ADN possible, ou LEF (Loop Extruding Factor) (ou plus généralement impliquée dans le bouclage de la chromatine, voir section 1.2.1), et divers travaux récents ont montré qu'il est fondamental de maintenir l'organisation génomique á l'échelle TAD : (Wutz et al., 2017)(Gassler et al., 2017)(Rao et al., 2017)(Haarhuis et al., 2017). Il est intéressant de noter que le m phénotype de TAD perturbés est observé lors de l'épuisement du CTCF (Nora et al., 2017). Les effets de l'appauvrissement en CTCF, en Cohesin ou en cofacteur de Cohesin sur les cartes Hi-C sont rapportés á la figure 1.13. En outre, les interactions á longue distance ont lieu de préférence aux sites de liaison convergents du CTCF et Rao et ses collègues ont montré que Cohesin se lie de préférence aux sites de liaison convergents du CTCF8 (Rao et al., 2014). L'ensemble de ces preuves prouve de manière frappante que CTCF et Cohesin sont impliqués dans la régulation de la structure de la chromatine au niveau de la TAD. On ne sait toujours pas comment une extrudeuse Cohesin peut fonctionner. Il n'est m pas clair si l'extrusion est effectuée par une Cohesin individuelle ou par deux ou plus. Le groupe de Cees Dekker a montré que la condensation in vitro n'extrude que d'un cté (Ganji et al., 2018). Il serait raisonnable de penser que Cohesin se comporterait de la m manière. Malgré l'absence de preuve directe de l'extrusion en boucle, il est clair que l'organisation chro-matin ne peut e le résultat d'un piégeage purement stochastique par Cohesin et il est encore moins probable

que le mécanisme d'entrament soit un pur mouvement thermique. Le besoin de sites convergents de la CTCF comme panneau STOP pour Cohesin, est une exigence forte. En effet, les altérations du motif CTCF, inversion ou suppression, conduisent á une perte du TAD correspondant (Lupi?an?ez et al., 2015).

Le but de ce travail est de décrire la dynamique de Cohesin et de déterminer la nature de l'interaction avec d'autres facteurs (CTCF, Sororin, Nipbl). Les principales découvertes dans le domaine de l'architecture de la chromatine et de ses régulateurs reposent sur des techniques statiques, basées sur la fixation cellulaire, et les conclusions sont tirées des moyennes de population. Notre intention était d'apporter notre contribution par une approche á une seule cellule, une seule molécule. En choisissant le suivi d'une seule molécule, nous avons pu étudier Cohesin dans l'espace et dans le temps, en ajoutant des informations sur sa dynamique.

Chapitre 3

L'objet de ce travail est la caractérisation de la dynamique des protéines nucléaires, dont le rle dans la régulation de la structure de la chromatine les amène á lier la chromatine. Comme une molécule liée reflète la diffusion sous-jacente de la chromatine, il est utile de rapporter la dynamique d'un locus pour avoir une référence. Par souci de cohérence avec les lignées cellulaires utilisées pour le suivi du CTCF et de la Cohésine, je me suis concentré sur la dynamique de la chromatine dans les cellules souches embryonnaires de souris (mESC). Il a en effet été démontré que la différenciation, ou plus précisément les changements dans le scénario transcriptionnel d'un tissu par rapport á une cellule polipotente, peut fortement affecter la mobilité de la chromatine (Gu et al., 2018). Les expériences présentées dans cette section ont été possibles gr á la générosité de la professeure Edith Heard et de son équipe, qui ont aimablement accepté de partager les lignées cellulaires. Je présente ici les résultats obtenus pour le Centre d'inactivation X (XIC) sur le chromosome X. La lignée cellulaire utilisée pour suivre le centre d'inactivation X (Xic), insérée via un réseau TetO et marquée avec eGFP, a été publiée dans (Masui et al., 2011) et (Giorgetti et al., 2016). Les valeurs du coefficient de diffusion apparent (D) sont limitées á une ré gion de très petites valeurs ($-3 < logD < -1$) et les exposants anomaux obtenus pour ces trajectoires sont significativement inférieurs á 1 ; pour ? les valeurs obtenues en ajustant le MSD sont très faibles, parfois trop proches de 0 ; cet effet peut e d á la longueur finie du MSD, car pour un objet qui ne se déplace pas trop, les premiers points sont dominés par l'erreur de localisation. La diffusion de la chromatine est clairement localisée dans un espace de paramètres de$\alpha < 0,5 et logD < -1$; ces valeurs seront utilisées comme référence pour identifier et valider la sous-population des molécules liées á l'ADN dans les sections

suivantes.

CTCF Pour caractériser la cinétique et la dynamique de liaison á la CTCF, j'ai suivi la protéine, couplée á un Halotag, dans des cellules souches embryonnaires de souris. La lignée cellulaire est représentée schématiquement dans le panneau D de 3.2 : le CTCF endogène est couplé avec l'Halotag pour effectuer le suivi d'une seule molécule. Les différents observables, c'est-á-dire la fraction de molécules liées, la probabilité de survie, la distribution des coefficients de diffusion, ont été extraits selon les méthodes décrites á la section 2.5 et tous les résultats sont résumés á la figure 3.2. Dans le panneau A, la fraction de CTCF lié et un contrle sont reportés. La moitié des molécules CTCF sont liées en S/G2, une valeur significativement plus élevée que le témoin. Comme déjá mentionné, la fraction de molécules liées est estimée á partir des acquisitions les plus rapides (5 ms d'exposition á 197 Hz) pour capturer les protéines les plus rapides. En ce qui concerne la répartition du temps de séjour, le CTCF montre une tendance du droit de l'électricité (panel B). Notre hypothèse est que de telles distributions découlent de la convolution de nombreux taux de dissociation liés á des événements de liaison transitoires. Il s'agit en effet d'un comportement récurrent de facteurs de transcription très différents á la recherche de leur cible (Lac répresseur : (Caccianini et al., 2015), Tet répresseur : (Nor- manno et al., 2015) et des données non publiées sur Transcription Activator Like Effectors (TALE)). Dans la fourchette de la fraction de secondes, nous échantillonnons très probablement des interactions non spécifiques. Les événements de liaison plus longs, de l'ordre de quelques centaines de secondes, sont considérés comme correspondant á des événements de liaison stables, très probablement á la séquence cible spécifique. Pour décrire la mobilité du CTCF, j'ai calculé le MSD pour chaque trajectoire et, en excluant le coefficient de diffusion apparent D, l'histogramme de log10(D) est présenté dans le panneau C de la figure 3.2. La distribution s'étend sur un large éventail de valeurs, depuis les valeurs très mobiles (entre 1 et 10 ?m2/s) jusqu'aux valeurs de D qui correspondent á la diffusion de la chromatine ($D$?0.01?$m2/s$). Deux populations sont visibles dans l'histogramme, correspondant á une sous-population de molécules liées (centrée autour de logD $\sim -2$) et une de molécules diffusantes (centrée sur$logD \sim 0$). Mais, pour les arguments exposés ci-dessus concernant le nombre de sous-populations dynamiques potentielles, je préfère présenter l'histogramme brut. Un moyen d'estimer la fraction liée des molécules consiste á examiner combien de valeurs de D sont inférieures au seuil dynamique imposé par la diffusion de la chromatine qui est logD ? -1 (Gu et al., 2018) et le travail personnel (voir section 3.1). Sur 915 trajectoires, 424 présentent un coefficient de diffusion comparable á celui de la chromatine, cohérent avec la valeur trouvée en regardant la distribution des longueurs de pas (panneau

A). Parallèlement á mes travaux, plusieurs études sur la trailité des mono-composants CTCF ont été publiées ; c'est pourquoi j'ai décidé de concentrer mes recherches sur Cohesin et de ne pas pousser plus loin l'analyse du comportement du CTCF. Les deux travaux (Hansen et al., 2017), (Agarwal et al., 2017) seront présentés par rapport á ce que j'ai fait dans la section 3.6.

Suivant l'approche décrite á la section 2.5, j'ai quantifié la dynamique de Rad21 dans des conditions de type sauvage (WT) dans la lignée cellulaire dont le génotype est montré á la figure 3.3. Comme le montre la figure 3.5 du panneau A, la fraction liée á Cohesin est d'environ 70 %, ce qui est nettement supérieur á la valeur trouvée pour le CTCF. Les cellules souches passent plus de 60% du cycle en phase S, comme le montrent (El-Badawy et El-Badri, 2016), et la coloration Hoechst nous a permis de confirmer que les cellules ne subissent pas la Mitose. Nous supposons que la valeur observée pour la fraction de molécules liées correspond aux cellules de S/G2, de sorte que le résultat est conforme á ce qui a été publié dans (Gerlich et al., 2006). La probabilité de survie montre une décroissance de la loi de puissance et atteint des valeurs plus élevées pour les événements de liaison les plus longs. Il est intéressant d'observer qu'une protéine qui n'a pas de domaine spécifique de liaison á l'ADN montre le m comportement qu'un facteur de transcription. Cette observation nous dit que Cohesin fait plein des interactions non-spécifiques. Il est intéressant de noter que la queue de la distribution atteint des temps de séjour plus élevés que le CTCF, en particulier qu'elle se décompose vers les 60 ans, juste avant le blanchiment. Des expériences FRAP ont montré que les liaisons stables de Cohesin peuvent durer jusqu'á 20 á 30 minutes (Gerlich et al., 2006) (Ladurner et al., 2014) (Hansen et al., 2017), mais avec une seule molécule, le suivi est difficile á atteindre dans ces délais. Le tracé présenté dans la figure 3.5 panel B est issu des ms conditions expérimentales que celles mentionnées pour le CTCF dans la section précédente (texp = 50ms en imagerie continue). Dans le panneau C, l'histogramme de logD est affiché. La distribution est très large, ce qui représente un scénario diffusif très hétérogène. Il y a un pic considérable autour de $logD \sim -1,75$, correspondant aux événements de liaison, et un autre pic autour de $logD \sim 1,5$, qui représente les molécules librement diffusives. Entre ces deux pics, il y a une région intermédiaire considérable, $-1 < logD < 1$ qui ne peut e facilement associée á une population diffuse spécifique. Contrairement á la majorité des fac-teurs nucléaires, en particulier les protéines de liaison á l'ADN, Cohesin se distingue par sa mobilité très hétérogène. La diversité de la dynamique de Cohesin peut e appréciée á partir du kymographe de la figure 3.4. Dans le panel D, j'ai rapporté la variabilité d'une cellule á l'autre pour les valeurs de logD : les distributions par cellules individuelles sont très similaires, donc les données

rapportées dans le panel C sont une représentation robuste du scénario de diffusivité en S/G2 pour Cohesin dans les cellules souches embryonnaires de souris. L'hétérogénéité du comportement diffusif de Cohesins est confirmée par des simulations réalisées par Simon Grosse-Holz, du groupe de Leonid Mirny (MIT, Cambridge, MA). Le pipeline de la simulation est le suivant : 1. générer une marche aléatoire avec un coefficient de diffusion D 2. ajouter l'erreur de localisation ? et le flou de mouvement 3. répéter, en générant un ensemble de trajectoires dont les longueurs et les temps d'exécution correspondent aux données réelles. Les valeurs de D et de ? utilisées pour les simulations sont estimées á partir de données hors périmétrie. Comme le montre la figure 3.6, une seule valeur de D n'est pas suffisante pour reproduire expérimentalement la distribution observée. Pour faire correspondre les données expérimentales et les données synthétiques, il est nécessaire de choisir non pas une valeur unique de D mais une distribution P(D), dans notre cas une distribution de Rayleigh. Nous concentrons actuellement nos travaux sur l'étude de la relation entre l'étalement des valeurs du coefficient de diffusion et l'encombrement éventuel.

Cohesine sans CTCF

L'un des objectifs de ce travail est d'étudier l'interaction entre les facteurs nucléaires impliqués dans la régulation de la structure tridimensionnelle de la chromatine. Pour atteindre cet objectif, nous avons décidé d'étudier d'abord la dynamique de Cohesin dans l'optique de la CTCF. Gr au travail d'Elph'ege Nora, dans le laboratoire de Benoit Bruneau au Gladstone Institute de San Francisco (USA), qui a réalisé la lignée cellulaire illustrée dans la figure 2.7, j'ai pu imaginer la sous-unité Rad21 de Cohesin dans les ms conditions que lors de ses travaux précédents (Nora et al, 2017) o il a montré que la diminution du CTCF entra une perte des TADs, de leur isolation et de leur positionnement aux sites CTCF. En effectuant le suivi d'une seule molécule dans des cellules vivantes, j'ai pu aborder le problème d'un point de vue dynamique et conserver l'information des molécules individuelles, dans des cellules uniques. Le génotype de la lignée cellulaire est esquissé á la figure 3.7. Rad21, ou Scc1, est un composant stable du complexe Cohesin ; il a été démontré par précipitation co-immunologique que la précipitation de Rad21 implique la pré-cipitation de tous les autres composants natifs de Cohesin (Smc1a, Smc3) (Hansen et al., 2017). Il est juste de supposer que lors de l'imagerie de Rad21, nous observons l'ensemble du complexe et j'utiliserai éventuellement Rad21 et Cohesin comme synonimes. Pour étudier Cohesin en l'absence de cellules CTCF ont été incubées pendant la nuit ($\sim 14 heures$) avec de l'auxine. Bien que la cinétique du système dégronique soit de l'ordre de quelques heures (voir fig. 2.8), nous avons choisi des temps d'incubation plus longs pour atteindre l'homogénéité de la population cellulaire, une sorte

d'état stable biologique. La suppression correcte a été vérifiée en fluorescence gr au rapporteur GFP couplé au CTCF (voir fig 3.8, panneaux centraux). Pour identifier les frontières nucléaires, surtout en l'absence de CTCF, j'ai coloré la chromatine avec 1?M de Hoechst (voir fig 3.8). Les expériences sur des cellules traitées et non traitées ont toujours été effectuées dans la m journée et dans les ms conditions d'imagerie. En cohérence avec la stratégie de suivi adoptée pour CTCF et Cohesin WT, j'ai réalisé une imagerie á vitesse d'acquisition rapide et extrapolé, á partir des ensembles de données relatifs, la fraction de molécules liées. Les résultats sont présentés dans la figure 3.9 panneau A, les cellules non traitées sont représentées en bleu et les cellules supplémentées en auxine sont en violet (le code de couleur est conservé sur l'ensemble du manuscrit). Il n'y a pas de différence significative dans la fraction de cohésines liées lors de l'épuisement du CTCF et les valeurs rapportées sont consistent avec ce qui a été trouvé pour Cohesin WT et avec la littérature (Gerlich et al., 2006) (Hansen et al., 2017) pour les cellules dans S/G2. Dans le panneau B de la m figure, j'ai indiqué que la probabilité de survie de Cohesin provenait de deux séries d'acquisitions, avec texp = 50ms et texp = 1s, toutes deux réalisées dans un régime d'imagerie en continu. Dans de telles conditions d'imagerie, j'ai pu acquérir des films qui ont duré 5000 images, ce qui correspond á $\sim 4minutes$. La courbe dérivée des expériences á l'échelle de temps la plus longue a été rééchelonnée avec la méthode illustrée dans la figure 2.14 et expliquée dans la section 2.5.2, visiblement le point de jonction des deux courbes se situe autour de 5 secondes. Dans le m tracé, la courbe de blanchiment est affichée, elle représente la désintégration du colorant pour texp = 50ms d'imagerie continue (avec les ms conditions d'imagerie de l'acquisition de données). La distribution des temps de séjour de Cohesin ne semble pas affectée par l'épuisement du CTCF á la minute près. Comme déjá mentionné, il a été démontré par différents groupes que Cohesin peut rester lié á la chromatine jusqu'á $\sim 20 - 30$ minutes et il est clair que je n'atteins pas cette échelle de temps avec mes expériences avec une seule molécule. Par conséquent, j'ai décidé de ne pas déduire les taux de désocialisation des distributions des probabilités de survie, car les valeurs résultantes ne décriraient que les sous-populations de Cohesin qui ont une liaison transitoire. J'aborderai le sujet plus en détail dans la section 3.6, mais globalement, l'absence de CTCF ne semble pas affecter la cinétique de liaison de Cohesin. D'autre part, la dynamique de Cohesin semble légèrement différente avec et sans CTCF. Dans les panneaux C et D de la figure 3.9, la distribution du coefficient de diffusion est indiquée. Les histogrammes du logD +/- CTCF de Cohesin ont la m distribution large, couvrant plus de 4 ordres de grandeur, indiquant qu'il n'y a aucun effet sur une sous-population diffuse particulière de Cohesin. Une augmentation des molécules mobiles

est visible dans l'histogramme et devient plus frappante lors du calcul de la fonction de distribution cumulative (FCD), montrée dans le panneau D. L'épuisement du FCCC implique une augmentation significative de la sous-population mobile. L'écart entre les deux CDF a été quantifié par un test de Kolmogorov-Smirnov qui a donné une valeur p de p = 0,0196, en supposant que les deux courbes expérimentales proviennent de la m distribution. Il est intéressant de noter que l'écart entre les deux CDF est localisé dans une plage de valeurs de logD qui couvre trois ordres de grandeur (-1 ¡ logD ¡ 1) et qu'il est le plus élevé dans une fene de logD qui est trop faible pour diffuser librement les protéines et trop élevée pour les molécules liées. Pour étudier plus en détail la fraction des cohésines affectées par l'épuisement du CTCF, j'ai isolé toutes les trajectoires se trouvant dans la région $-1,5 < logD < 0,2$ et j'ai tracé la distribution de la longueur des pas, les résultats sont indiqués en haut de la figure 3.10. Il n'y a pas de différence frappante entre les deux distributions et, sur le plan cinétique, une seule population est observée dans les deux cas. Pour compléter l'étude sur la dynamique de Cohesin avec et sans CTCF, j'ai extrait l'exposant anomal ? de chaque trajectoire, comme expliqué en 2.5.3, et le tracer en fonction du D de la m trajectoire dans les nuages de points de la figure 3.10. Un examen très attentif est nécessaire pour apprécier les légères différences entre les deux parcelles : il y a moins de cohortes immobiles en l'absence de CTCF. Dans l'ensemble, la cinétique de liaison ne semble pas e affectée par l'absence de CTCF car la fraction de molécules liées reste inchangée et la distribution du temps de séjour est constante jusqu'á l'échelle de temps infime. Ces deux conclusions sont cohérentes avec les résultats ChIP-Seq de la figure 3.27 : la quantité de Cohesin sur la chromatine n'est pas modifiée en l'absence de CTCF. La dynamique de Cohesin est légèrement affectée par le knock out du CTCF, mais la nature de la perturbation n'est pas encore claire. Il n'y a pas de variation significative dans les espèces dynamiques observées pour Cohesin. Cohesin dans des céllules non-cycling

Le comportement de la cohésine est étroitement lié au cycle cellulaire, comme décrit brièvement á la section 1.3.2 et á la figure 3.11. En particulier au début de la phase S, une sous-population de Cohésines est acétylée et verrouillée sur la chromatine pour assurer une cohésion chromatidique adéquate des chromatides soeurs (Peters et al., 2008). Cette étude est centrée sur le comportement de Cohesin dans la régulation de l'organisation spatiale des chromosomes interphasiques, d'o la nécessité de discerner les cohésines cohésives, o Cohesin cohésif indique un complexe de Cohesin impliqué dans la cohésion des chromatides soeurs, acétylée et stabilisée par Sororin. Suivant l'approche adoptée par Elph'ege Nora dans son dernier travail, j'ai par- formé les expériences de suivi dans des cellules non cycliques. D'autres travaux se

sont appuyés sur des marqueurs de cycle comme PCNA (Gerlich et al., 2006), le système Fucci (Hansen et al., 2017) mais ces approches n'excluent pas les Cohesins cohésifs du tableau. De plus, dans les études citées, les auteurs ont réalisé des expériences de récupération de fluorescence après photolixiviation (FRAP) sur Cohesin dont les résultats sont des valeurs moyennes sur la mobilité de l'ensemble des molécules. Choisir des cellules non cycliques nous a donné la certitude que chaque Cohesin que nous observions par le suivi d'une seule molécule n'était pas impliquée dans la cohésion chromatique-matière sur. Conformément aux observations publiées dans (Nora et al., 2017), nous avons remarqué que les cellules Rad21-Halotag du CTCF-AID ont cessé de répondre á l'auxine lors de la différenciation, probablement en raison de la réduction au silence du transgène Tir1 intégré au dme de course. Nous avons surmonté ce problème en créant une autre lignée cellulaire CTCF-AID avec le transgène Tir1 ciblant le locus Tigre, qui est resté stable lors de la différenciation cellulaire en progéniteurs neuraux et en astrocytes, comme le montre la figure 3.12. Comme pour le CTCF et Cohesin, j'ai réalisé des expériences de suivi de molécules uniques avec l'imagerie continue HiLO á texp = 5ms, pour quantifier la fraction de molécules liées, á texp = 50ms, pour caractériser la dynamique á ex-tract la distribution des temps de séjour. Comme indiqué précédemment, les astrocytes sont différenciés des cellules souches utilisées pour étudier la Cohésine en présence et en l'absence de CTCF, ce qui m'a permis d'étudier le comportement de la Cohésine +/- CTCF en G0, c'est-á-dire en l'absence de Cohésines cohésives. Les résultats sont présentés á la figure 3.13. 50% des cohésines sont liées en G0 et il n'y a pas de différence de présence ou d'absence de CTCF (fig 3.13 panneau A). La distribution des temps de résidence, présentée dans le panel B comme la probabilité de survie, suit la m décroissance que les courbes précédemment présentées pour Cohesin (les résultats sont fusionnés dans la figure 3.28). Il est important de noter que dans ce cas, la différence entre les deux CDF de la logD est encore plus importante que dans le cas de l'ES. Un test de Kolmogorov-Smirnov a donné une valeur p de p = 0,0014, l'hypothèse nulle étant que les deux courbes appartiennent á la m distribution. Cette légère différence peut également e appréciée dans le nuage de points de la figure 3.14 : en traitant les cellules avec de l'auxine, on observe une légère augmentation de la fraction mobile ($\alpha \sim 1 etlogD \sim 0$). Nous avons ensuite cherché á corroborer les résultats dans des cellules non cycliques avec des ex-périmentations abrogeant la chromatide soeur Cohesin dans des cellules ES.

Autres mutants

Cohesine sans Sororine Dans cette section, je présenterai les résultats des expériences de suivi sur Cohesin en l'absence de Sororin. Avant d'entrer dans le détail des résultats, je présenterai brièvement Sororin et son rle par rap-

port au complexe Cohesin. La sororine est une protéine vertébrée nécessaire pour fixer la Cohésine sur la chromatine et assurer la cohésion des chromatides soeurs. La sororine stabilise la liaison Cohesin-ADN en masquant le domaine reconnu par Wapl, Déchargeur de Cohesin. En fait, la dissociation acétylée de la Cohésine peut e réalisée une fois que la Sororine est phosphorée et déchargée du complexe Cohesin, laissant la place á Wapl pour ouvrir l'anneau et libérer Cohesin (Nishiyama et al., 2013). La sororine est recrutée sur chromatine déjá en phase S (Nishiyama et al., 2010) et peu après la réplication de l'ADN (Lafont et al., 2010) et en l'absence du complexe Cohesin, elle ne peut e chargée. La liaison de la sororine au complexe Cohesin est liée á l'acétylation de la cohésine : deux groupes différents ont montré que les acétyltransférases Esco1/Esco2 sont nécessaires, mais pas suffisantes, pour la liaison Sororin- Cohesin (Nishiyama et al., 2010) (Lafont et al., 2010). La désactivation de la sororine affecte considérablement la cohésion des chromatides soeurs et les cellules manquant de sororine finissent par se bloquer dans la mitose (Rankin, 2005), comme l'illustre la figure 3.15. Suite á la stratégie d'épuisement utilisée pour le CTCF (voir section 3.4), la sororine a été couplée au système de dégradation inductible par l'auxine dans une lignée cellulaire Cohesin-Halo (2.3). La cinétique de la déplétion en sororine lors de l'incubation avec l'auxine est montrée á la figure 3.16. J'ai étudié le comportement de Cohesin en l'absence de Sororin avec deux temps d'incubation différents : après 3 heures et 6 heures d'incubation de l'auxine. Les cellules imagées après 3 heures d'incubation avec l'auxine étaient pour la plupart encore en S/G2 (environ 1 cellule sur 20 avait un phénotype mitotique) voir fig 3.17 pour un exemple. Le temps d'incubation choisi est un rapport fin entre le temps nécessaire pour obtenir une dégradation homogène de la protéine (voir fig 2.8 pour la cinétique du dégron) tout en évitant une incubation plus longue qui induit inévitablement la majorité des cellules de la Mitose. Afin d'exclure les cellules qui se seraient bloquées prématurément dans la mitose, et de pouvoir faire une comparaison directe avec les données précédentes collectées en interphase, et en cohérence avec le protocole adopté pour les expériences présentées dans les chapitres précédents, les cellules ont été colorées avec Hoechst. Dans la figure 3.17, un exemple des cellules S/G2 imagées est présenté, tandis que la figure 3.18 montre certaines des images du signal de Hoechst dans les cellules mitotiques. Les résultats du suivi de la Cohésine en l'absence de Sororine tant pour les cellules en S/G2 que pour celles en Mitose sont montrés dans la figure 3.19. Le panel A rapporte la quantification de la fraction de molécules liées de Cohesin en l'absence de Sororin dans les cellules S/G2 (violet) et dans Mitosis (émeraude). L'absence de Sororine affecte légèrement la fraction de Cohésines liées tandis que dans la Mitose la portion de molécules liées est encore plus faible que dans le

contrle (violet). Le faible effet dans les cellules S/G2 peut e d au fait que la sororine interagit avec les co-résines qui sont déjá chargées topologiquement sur la chromatine et acétylées (Nishiyama et al., 2010) (Lafont et al., 2010) et que cette sous-population de co-résines est peu importante ($\sim 10\%$ selon nos expériences). D'autre part, pendant la prophase, les anneaux de Cohesin sont ouverts par des sépararases, laissant la place aux Condensins : Les cohésines sont libres de se déplacer dans un espace plus large et moins encombré car l'ADN est compacté dans les chromosomes mitotiques (voir figure 3.18) (Nasmyth, 2001). Le gain dans l'espace accessible est très probablement la raison principale derrière la plus petite fraction de molécules liées dans la Mitose par rapport au contrle. Dans le panneau B de la m figure, le lecteur peut apprécier les deux probabilités de survie de Cohesin dans les cellules dépourvues de sororine. Déjá dans S/G2, la probabilité de survie chute plus rapidement que dans les exemples présentés précédemment (voir fig. 3.9 et fig. 3.13 pour comparaison). Si l'on s'en tient á l'idée que seules les co-résines acétylées sont stabilisées par la sororine, ce résultat indique que les cohésines cohésives sont en fait une sous-population que nous capturons avec notre imagerie. C'est peut-e aussi le signe que malgré l'arru cycle cellulaire induit chez les Astrocytes, des acteurs comme Sororin, qui sont cruciaux pour la cohé- sion chromatide sur, sont toujours présents et accomplissent leur te. La distribution des probabilités de survie pour Cohesin dans Mitosis confirme le scénario d'une diffusion libre des Cohesin puisque l'enregistreur de temps d'association est de l'ordre des fractions de secondes. Les histogrammes du logarithme du coefficient de diffusion apparent sont représentés dans les panneaux C et D de la figure 3.19. Il est intéressant de noter qu'en cas d'épuisement des sororines en phase S/G2, la cohésine perd non seulement plus de la moitié des molécules liées de fa stable (c'est-á-dire $-2 < logD < 1$), mais également la population á diffusion libre (c'est-á-dire $0 < logD < 1$). Ce résultat controversé n'est pas facile á interpréter dans le contexte décrit ; une enqu et une réflexion plus approfondies sont nécessaires. Néanmoins, l'histogramme de Cohesin in Mitosis renforce le tableau dressé avec les autres résultats : la grande majorité des molécules présentent un coefficient de diffusion élevé (i.e. $0 < logD < 1$).

Cohesine sans Nipbl

Cette dernière section des résultats, est consacrée á l'investigation des dy-namics de Cohesin en l'absence de Nipbl (Scc2). Suivant l'approche du paragraphe précédent, je décrirai d'abord Nipbl et son rle dans le contexte du complexe de Co-Hesin, puis j'illustrerai les observations. Nipbl fait partie du complexe de chargement de Cohesin qui consiste en Scc2 (Nipbl) et Scc4 (Mau2). Le Nipbl est indispensable pour charger le complexe sur la chromatine, mais il n'est pas nécessaire pour la cohésion de la chromatine soeur

(Ciosk et al., 2000) (Mu- Rayama et Uhlmann, 2014). Le Nipbl est également connu pour déclencher la capacité de Cohesin á hydrolyser l'ATP, et il a été récemment démontré que le Nipbl est nécessaire et suffisant pour stimuler l'activité ATPase de Cohesin en présence de l'ADN (Petela et al., 2018). Comme pour les autres mutants, l'épuisement du Nipbl a été obtenu par le système dégronique décrit á la section 2.4.1. L'appauvrissement en Nibpl a des effets plus dramatiques sur la viabilité cellulaire étonnamment, la fraction des cohésines liées n'est pas significativement affectée par l'épuisement du Nipbl. Comme le Nipbl est nécessaire pour la charge topologique de Cohesin sur la chromatine, ces résultats pourraient indiquer que nous sommes en train de déconnecter des interactions non spécifiques Cohesin-ADN, c'est-á-dire des interactions qui ne nécessitent pas de charge topologique et qui ne sont pas liées á une extrusion en boucle ou á une cohésion chromatidique sur. Cependant, la probabilité de survie de Cohesin diminue de fa dramatique en l'absence de Nibpl, ce qui contredit l'idée d'interactions purement non spécifiques. Il est possible que la fréquence d'image élevée choisie pour quantifier la fraction de molécules liées (c.-á-d. 197 Hz) implique un suréchantillonnage des interactions transitoires. La distribution de logD est également assez surprenante puisque nous observons la perte d'une partie considérable des cohésines á diffusion libre, celles qui ont $logD > 0$, ce qui est franchement un résultat difficile á interpréter et d'autres expériences ainsi que d'autres outils de quantification sont très probablement nécessaires pour mieux caractériser le comportement de Cohesin en l'absence du Nipbl.

Control

Notre expérience de contrle consiste á suivre le Halotag, avec le colorant organique JF549, couplé á un signal de localisation nucléaire (NLS). Les expériences ont été réalisées sur la lignée cellulaire montrée en 3.23, panel A. Le Halotag-NLS est codé dans le locus CTCF avec un peptide clivant 2A (Wang et al., 2015). Le clivage se produit pendant la traduction lorsque, dans notre cas, le Halotag-NLS est coupé du CTCF. Le panneau B de la figure 3.23 présente la distribution des probabilités de survie pour la control, qui se dégrade plus rapidement que toute autre survie présentée dans ce manuscrit, á l'exception du suivi dans les cellules mitotiques. L'histogramme des coefficients de diffusion montre encore une large distribution qui s'étend á des valeurs très faibles qui peuvent e un symptme d'interactions transitoires (panneau C). D'autres groupes ont signalé que le Halo-NLS peut se lier et/ou interagir de fa transitoire (un exemple en est donné par Hansen et al., 2017). Malgré ces interactions très brèves, il est possible d'établir une distinction claire entre ces événements rapides et transitoires et ceux qui sont provoqués par une protéine comme l'Halo-NLS. Le suivi des Halo-NLS dans d'autres systèmes (suivi des protéines TALE dans U2OS, travail

personnel) a montré une dégradation très rapide á l'échelle des fractions de secondes. Il est possible que dans le contrle utilisé pour ce travail, le système de clivage ne soit pas efficace á 100% et que les quelques interactions plus longues, á l'échelle de la seconde, soient réalisées par une version tronquée du CTCF. En ce sens, le suivi de Cohesin dans les cellules mitotiques offre un bon contrle pour l'identification de la sous-population á diffusion libre. D'autre part, les résultats dans les cellules mitotiques reflètent une situation très particulière o la chromatine, l'élément le plus encombré dans le noyau, est fortement condensée, laissant plus d'espace á la protéine pour diffuser. Comme la perspective de la caractérisation de Cohesin déplacera l'échelle de temps d'intérers la dizaine de minutes, je crois que les contrles présentés sont fiables.

# Bibliography

H. Agarwal, M. Reisser, C. Wortmann, and J. C. M. Gebhardt. Direct Observation of Cell-Cycle-Dependent Interactions between CTCF and Chromatin. *Biophysical Journal*, 112(10):2051–2055, May 2017. ISSN 00063495. doi: 10.1016/j.bpj.2017.04.018. URL https://linkinghub.elsevier.com/retrieve/pii/S0006349517304344.

E. Alipour and J. F. Marko. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Research*, 40(22):11202–11212, Dec. 2012. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gks925. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks925.

D. E. Anderson, A. Losada, H. P. Erickson, and T. Hirano. Condensin and cohesin display different arm conformations with characteristic hinge angles. *The Journal of Cell Biology*, 156(3):419–424, Feb. 2002. ISSN 0021-9525, 1540-8140. doi: 10.1083/jcb.200111002. URL http://www.jcb.org/lookup/doi/10.1083/jcb.200111002.

M. Andrade and P. Bork. HEAT repeats in the Huntington's disease protein. *HEAT repeats in the Huntington's disease protein.*, 11(2):115–6, 1995. doi: 10.1038/ng1095-115.

B. J. Beliveau, E. F. Joyce, N. Apostolopoulos, F. Yilmaz, C. Y. Fonseka, R. B. McCole, Y. Chang, J. B. Li, T. N. Senaratne, B. R. Williams, J.-M. Rouillard, and C.-t. Wu. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proceedings of the National Academy of Sciences*, 109(52):21301–21306, Dec. 2012. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1213818110. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1213818110.

F. Benedetti, J. Dorier, Y. Burnier, and A. Stasiak. Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. *Nucleic Acids Research*, 42(5):2848–2855, Mar.

2014. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkt1353. URL https://academic.oup.com/nar/article/42/5/2848/1063457.

W. A. Bickmore. The Spatial Organization of the Human Genome. *Annual Review of Genomics and Human Genetics*, 14(1):67–84, Aug. 2013. ISSN 1527-8204, 1545-293X. doi: 10.1146/annurev-genom-091212-153515. URL http://www.annualreviews.org/doi/10.1146/annurev-genom-091212-153515.

B. Bintu, L. J. Mateo, J.-H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, and X. Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, Oct. 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aau1783. URL http://www.sciencemag.org/lookup/doi/10.1126/science.aau1783.

A. N. Boettiger, B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, G. Fudenberg, M. Imakaev, L. A. Mirny, C.-t. Wu, and X. Zhuang. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586):418–422, Jan. 2016. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature16496. URL http://www.nature.com/articles/nature16496.

R. A. Bradshaw and P. D. Stahl. *Encyclopedia Of Cell Biology*. Elsevier edition, 2016.

K. Bystricky. Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. *FEBS Letters*, 589(20PartA):3014–3022, Oct. 2015. ISSN 00145793. doi: 10.1016/j.febslet.2015.07.012. URL http://doi.wiley.com/10.1016/j.febslet.2015.07.012.

L. Caccianini, D. Normanno, I. Izeddin, and M. Dahan. Single molecule study of non-specific binding kinetics of LacI in mammalian cells. *Faraday Discussions*, 184:393–400, 2015. ISSN 1359-6640, 1364-5498. doi: 10.1039/C5FD00112A. URL http://xlink.rsc.org/?DOI=C5FD00112A.

D. I. Cattoni, A. M. Cardozo Gizzi, M. Georgieva, M. Di Stefano, A. Valeri, D. Chamousset, C. Houbron, S. Déjardin, J.-B. Fiche, I. González, J.-M. Chang, T. Sexton, M. A. Marti-Renom, F. Bantignies, G. Cavalli, and M. Nollmann. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nature Communications*, 8(1):1753, Dec. 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01962-x. URL http://www.nature.com/articles/s41467-017-01962-x.

C. Chapard, R. Jones, T. van Oepen, J. C. Scheinost, and K. Nasmyth. Sister DNA Entrapment between Juxtaposed Smc Heads and Kleisin of the Cohesin Complex. *Molecular Cell*, page S1097276519303958, June 2019. ISSN 10972765. doi: 10.1016/j.molcel.2019.05.023. URL https://linkinghub.elsevier.com/retrieve/pii/S1097276519303958.

J. Chaumeil. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes & Development*, 20(16):2223–2237, Aug. 2006. ISSN 0890-9369. doi: 10.1101/gad.380906. URL http://www.genesdev.org/cgi/doi/10.1101/gad.380906.

R. Ciosk, M. Shirayama, A. Shevchenko, T. Tanaka, A. Toth, A. Shevchenko, and K. Nasmyth. Cohesin's Binding to Chromosomes Depends on a Separate Complex Consisting of Scc2 and Scc4 Proteins. *Molecular Cell*, 5 (2):243–254, Feb. 2000. ISSN 10972765. doi: 10.1016/S1097-2765(00)80420-7. URL https://linkinghub.elsevier.com/retrieve/pii/S1097276500804207.

T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, 2(4):292–301, Apr. 2001. ISSN 1471-0056, 1471-0064. doi: 10.1038/35066075. URL http://www.nature.com/articles/35066075.

I. F. Davidson, D. Goetz, M. P. Zaczek, M. I. Molodtsov, P. J. Huis in 't Veld, F. Weissmann, G. Litos, D. A. Cisneros, M. Ocampo-Hafalla, R. Ladurner, F. Uhlmann, A. Vaziri, and J. Peters. Rapid movement and transcriptional re-localization of human cohesin on DNA. *The EMBO Journal*, 35 (24):2671–2685, Dec. 2016. ISSN 0261-4189, 1460-2075, 0261-4189, 1460-2075. doi: 10.15252/embj.201695402. URL http://emboj.embopress.org/lookup/doi/10.15252/embj.201695402.

J. Dekker. Capturing Chromosome Conformation. *Science*, 295(5558):1306–1311, Feb. 2002. ISSN 00368075, 10959203. doi: 10.1126/science.1067799. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1067799.

J. Dekker. Mapping the 3d genome: Aiming for consilience. *Nature Reviews Molecular Cell Biology*, 17(12):741–742, Dec. 2016. ISSN 1471-0072, 1471-0080. doi: 10.1038/nrm.2016.151. URL http://www.nature.com/articles/nrm.2016.151.

J. Dekker and L. Mirny. The 3d Genome as Moderator of Chromosomal Communication. *Cell*, 164(6):1110–1121, Mar. 2016. ISSN 00928674. doi: 10.1016/j.cell.2016.02.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867416300733.

E. de Wit, E. Vos, S. Holwerda, C. Valdes-Quezada, M. Verstegen, H. Teunissen, E. Splinter, P. Wijchers, P. Krijger, and W. de Laat. CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell*, 60(4):676–684, Nov. 2015. ISSN 10972765. doi: 10.1016/j.molcel.2015.09.023. URL https://linkinghub.elsevier.com/retrieve/pii/S1097276515007625.

J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, May 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11082. URL http://www.nature.com/articles/nature11082.

E. Dolgin. One of the most puzzling problems in genome architecture may have a simple solution. But no one can agree on what powers the process. page 3.

A. D. Edelstein, M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, and N. Stuurman. Advanced methods of microscope control using Manager software. *Journal of Biological Methods*, 1(2):10, Nov. 2014. ISSN 2326-9901. doi: 10.14440/jbm.2014.36. URL http://www.jbmethods.org/jbm/article/view/36.

A. El-Badawy and N. El-Badri. The cell cycle as a brake for -cell regeneration from embryonic stem cells. *Stem Cell Research & Therapy*, 7(1):9, Dec. 2016. ISSN 1757-6512. doi: 10.1186/s13287-015-0274-z. URL http://stemcellres.com/content/7/1/9.

F. Etoc, E. Balloul, C. Vicario, D. Normanno, D. Liße, A. Sittner, J. Piehler, M. Dahan, and M. Coppey. Non-specific interactions govern cytosolic diffusion of nanosized objects in mammalian cells. *Nature Materials*, 17(8):740–746, Aug. 2018. ISSN 1476-1122, 1476-4660. doi: 10.1038/s41563-018-0120-7. URL http://www.nature.com/articles/s41563-018-0120-7.

M. Falk, Y. Feodorova, N. Naumova, M. Imakaev, B. R. Lajoie, H. Leonhardt, B. Joffe, J. Dekker, G. Fudenberg, I. Solovei, and L. Mirny. Heterochromatin drives organization of conventional and inverted nuclei. preprint,

Biophysics, Jan. 2018. URL http://biorxiv.org/lookup/doi/10.1101/244038.

I. M. Flyamer, J. Gassler, M. Imakaev, H. B. Brandão, S. V. Ulianov, N. Abdennur, S. V. Razin, L. A. Mirny, and K. Tachibana-Konwalski. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*, 544(7648):110–114, Apr. 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature21711. URL http://www.nature.com/articles/nature21711.

G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. Mirny. Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, 15(9):2038–2049, May 2016. ISSN 22111247. doi: 10.1016/j.celrep.2016.04.085. URL https://linkinghub.elsevier.com/retrieve/pii/S2211124716305307.

G. Fudenberg, N. Abdennur, M. Imakaev, A. Goloborodko, and L. A. Mirny. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harbor Symposia on Quantitative Biology*, 82:45–55, 2017. ISSN 0091-7451, 1943-4456. doi: 10.1101/sqb.2017.82.034710. URL http://symposium.cshlp.org/lookup/doi/10.1101/sqb.2017.82.034710.

M. Ganji, I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, and C. Dekker. Real-time imaging of DNA loop extrusion by condensin. *Science*, 360(6384):102–105, Apr. 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aar7831. URL http://www.sciencemag.org/lookup/doi/10.1126/science.aar7831.

J. Gassler, H. B. Brandão, M. Imakaev, I. M. Flyamer, S. Ladstätter, W. A. Bickmore, J. Peters, L. A. Mirny, and K. Tachibana. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *The EMBO Journal*, 36(24):3600–3618, Dec. 2017. ISSN 0261-4189, 1460-2075, 0261-4189, 1460-2075. doi: 10.15252/embj.201798083. URL http://emboj.embopress.org/lookup/doi/10.15252/embj.201798083.

D. Gerlich, B. Koch, F. Dupeux, J.-M. Peters, and J. Ellenberg. Live-Cell Imaging Reveals a Stable Cohesin-Chromatin Interaction after but Not before DNA Replication. *Current Biology*, 16(15):1571–1578, Aug. 2006. ISSN 09609822. doi: 10.1016/j.cub.2006.06.068. URL https://linkinghub.elsevier.com/retrieve/pii/S0960982206018471.

T. Germier, S. Audibert, S. Kocanova, D. Lane, and K. Bystricky. Real-time imaging of specific genomic loci in eukaryotic cells using the ANCHOR

DNA labelling system. *Methods*, 142:16–23, June 2018. ISSN 10462023. doi: 10.1016/j.ymeth.2018.04.008. URL https://linkinghub.elsevier.com/retrieve/pii/S1046202317302335.

J. H. Gibcus, K. Samejima, A. Goloborodko, I. Samejima, N. Naumova, J. Nuebler, M. T. Kanemaki, L. Xie, J. R. Paulson, W. C. Earnshaw, L. A. Mirny, and J. Dekker. A pathway for mitotic chromosome formation. *Science*, 359(6376):eaao6135, Feb. 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aao6135. URL http://www.sciencemag.org/lookup/doi/10.1126/science.aao6135.

L. Giorgetti, R. Galupa, E. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard. Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription. *Cell*, 157(4):950–963, May 2014. ISSN 00928674. doi: 10.1016/j.cell.2014.03.025. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867414003614.

L. Giorgetti, B. R. Lajoie, A. C. Carter, M. Attia, Y. Zhan, J. Xu, C. J. Chen, N. Kaplan, H. Y. Chang, E. Heard, and J. Dekker. Structural organization of the inactive X chromosome in the mouse. *Nature*, 535(7613):575–579, July 2016. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature18589. URL http://www.nature.com/doifinder/10.1038/nature18589.

T. Gligoris and J. Löwe. Structural Insights into Ring Formation of Cohesin and Related Smc Complexes. *Trends in Cell Biology*, 26(9):680–693, Sept. 2016. ISSN 09628924. doi: 10.1016/j.tcb.2016.04.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0962892416300149.

J. B. Grimm, B. P. English, J. Chen, J. P. Slaughter, Z. Zhang, A. Revyakin, R. Patel, J. J. Macklin, D. Normanno, R. H. Singer, T. Lionnet, and L. D. Lavis. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods*, 12(3):244–250, Mar. 2015. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.3256. URL http://www.nature.com/articles/nmeth.3256.

B. Gu, T. Swigut, A. Spencley, M. R. Bauer, M. Chung, T. Meyer, and J. Wysocka. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science*, 359(6379):1050–1055, Mar. 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aao3136. URL http://www.sciencemag.org/lookup/doi/10.1126/science.aao3136.

L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van

Steensel. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197):948–951, June 2008. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature06947. URL http://www.nature.com/articles/nature06947.

J. H. Haarhuis, R. H. van der Weide, V. A. Blomen, J. O. Yáñez-Cuna, M. Amendola, M. S. van Ruiten, P. H. Krijger, H. Teunissen, R. H. Medema, B. van Steensel, T. R. Brummelkamp, E. de Wit, and B. D. Rowland. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*, 169(4):693–707.e14, May 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.04.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417304269.

B. Hajj, M. El Beheiry, I. Izeddin, X. Darzacq, and M. Dahan. Accessing the third dimension in localization-based super-resolution microscopy. *Phys. Chem. Chem. Phys.*, 16(31):16340–16348, 2014. ISSN 1463-9076, 1463-9084. doi: 10.1039/C4CP01380H. URL http://xlink.rsc.org/?DOI=C4CP01380H.

A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, and X. Darzacq. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. page 33, 2017.

A. S. Hansen, M. Woringer, J. B. Grimm, L. D. Lavis, R. Tjian, and X. Darzacq. Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife*, 7, Jan. 2018. ISSN 2050-084X. doi: 10.7554/eLife.33125. URL https://elifesciences.org/articles/33125.

M. V. Imakaev, G. Fudenberg, and L. A. Mirny. Modeling chromosomes: Beyond pretty pictures. *FEBS Letters*, 589(20PartA):3031–3036, Oct. 2015. ISSN 00145793. doi: 10.1016/j.febslet.2015.09.004. URL http://doi.wiley.com/10.1016/j.febslet.2015.09.004.

D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*, 42(15):9553–9561, Sept. 2014. ISSN 1362-4962, 0305-1048. doi: 10.1093/nar/gku698. URL http://academic.oup.com/nar/article/42/15/9553/2436727/Modeling-epigenome-folding-formation-and-dynamics.

M. H. Kagey, J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, D. J. Taatjes, J. Dekker, and R. A. Young. Mediator and cohesin connect gene

expression and chromatin architecture. *Nature*, 467(7314):430–435, Sept. 2010. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature09380. URL http://www.nature.com/articles/nature09380.

M. Kanke, E. Tahara, P. J. Huis in't Veld, and T. Nishiyama. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *The EMBO Journal*, 35(24):2686–2698, Dec. 2016. ISSN 0261-4189, 1460-2075, 0261-4189, 1460-2075. doi: 10.15252/embj.201695756. URL http://emboj.embopress.org/lookup/doi/10.15252/embj.201695756.

T. H. Kim, Z. K. Abdullaev, A. D. Smith, K. A. Ching, D. I. Loukinov, R. Green, M. Q. Zhang, V. V. Lobanenkov, and B. Ren. Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell*, 128(6):1231–1245, Mar. 2007. ISSN 00928674. doi: 10.1016/j.cell.2006.12.048. URL https://linkinghub.elsevier.com/retrieve/pii/S009286740700205X.

J. Krefting, M. A. Andrade-Navarro, and J. Ibn-Salem. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biology*, 16(1):87, Dec. 2018. ISSN 1741-7007. doi: 10.1186/s12915-018-0556-x. URL https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-018-0556-x.

R. Ladurner, V. Bhaskara, P. Huis in 't Veld, I. Davidson, E. Kreidl, G. Petzold, and J.-M. Peters. Cohesin's ATPase Activity Couples Cohesin Loading onto DNA with Smc3 Acetylation. *Current Biology*, 24(19):2228–2237, Oct. 2014. ISSN 09609822. doi: 10.1016/j.cub.2014.08.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0960982214009890.

A. L. Lafont, J. Song, and S. Rankin. Sororin cooperates with the acetyltransferase Eco2 to ensure DNA replication-dependent sister chromatid cohesion. *Proceedings of the National Academy of Sciences*, 107(47):20364–20369, Nov. 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1011069107. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1011069107.

P. R. Langer-Safer, M. Levine, and D. C. Ward. Immunological method for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy of Sciences*, 79(14):4381–4385, July 1982. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.79.14.4381. URL http://www.pnas.org/cgi/doi/10.1073/pnas.79.14.4381.

T. B. K. Le, M. V. Imakaev, L. A. Mirny, and M. T. Laub. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science*, 342(6159):731–734, Nov. 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1242059. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1242059.

E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326(5950):289–293, Oct. 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1181369. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1181369.

A. Losada. Cohesin in cancer: chromosome segregation and beyond. *Nature Reviews Cancer*, 14(6):389–393, June 2014. ISSN 1474-175X, 1474-1768. doi: 10.1038/nrc3743. URL http://www.nature.com/articles/nrc3743.

D. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, 161(5):1012–1025, May 2015. ISSN 00928674. doi: 10.1016/j.cell.2015.04.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867415003773.

M. Marbouty, A. Le Gall, D. Cattoni, A. Cournac, A. Koh, J.-B. Fiche, J. Mozziconacci, H. Murray, R. Koszul, and M. Nollmann. Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Molecular Cell*, 59(4):588–602, Aug. 2015. ISSN 10972765. doi: 10.1016/j.molcel.2015.07.020. URL https://linkinghub.elsevier.com/retrieve/pii/S109727651500578X.

J. F. Marko and E. D. Siggia. Polymer Models of Meiotic and Mitotic Chromosomes. *Molecular Biology of the Cell*, 8(11):2217–2231, Nov. 1997. ISSN 1059-1524, 1939-4586. doi: 10.1091/mbc.8.11.2217. URL http://www.molbiolcell.org/doi/10.1091/mbc.8.11.2217.

J. F. Marko, P. De Los Rios, A. Barducci, and S. Gruber. DNA-segment-capture model for loop extrusion by structural maintenance of chromosome (SMC) protein complexes. preprint, Biophysics, May 2018. URL http://biorxiv.org/lookup/doi/10.1101/325373.

O. Masui, I. Bonnet, P. Le Baccon, I. Brito, T. Pollex, N. Murphy, P. Hupé, E. Barillot, A. Belmont, and E. Heard. Live-Cell Chromosome Dynamics and Outcome of X Chromosome Pairing Events during ES Cell Differentiation. *Cell*, 145(3):447–458, Apr. 2011. ISSN 00928674. doi: 10.1016/j.cell.2011.03.032. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867411003096.

D. Mazza, A. Abernathy, N. Golob, T. Morisaki, and J. G. McNally. A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Research*, 40(15):e119–e119, Aug. 2012. ISSN 1362-4962, 0305-1048. doi: 10.1093/nar/gks701. URL https://academic.oup.com/nar/article/40/15/e119/1225229.

M. Merkenschlager and E. P. Nora. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annual Review of Genomics and Human Genetics*, 17(1):17–43, Aug. 2016. ISSN 1527-8204, 1545-293X. doi: 10.1146/annurev-genom-083115-022339. URL http://www.annualreviews.org/doi/10.1146/annurev-genom-083115-022339.

A. Monneron and W. Bernhard. Fine structural organization of the interphase nucleus in some mammalian cells. *Journal of Ultrastructure Research*, 27(3-4):266–288, May 1969. ISSN 00225320. doi: 10.1016/S0022-5320(69)80017-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0022532069800171.

Y. Murayama and F. Uhlmann. Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature*, 505(7483):367–371, Jan. 2014. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature12867. URL http://www.nature.com/articles/nature12867.

Y. Murayama and F. Uhlmann. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell*, 163(7):1628–1640, Dec. 2015. ISSN 00928674. doi: 10.1016/j.cell.2015.11.030. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867415015494.

T. Nagano, Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, and P. Fraser. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64, Oct.

2013. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature12593. URL http://www.nature.com/articles/nature12593.

K. Nasmyth. Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis. *Annual Review of Genetics*, 35(1):673–745, Dec. 2001. ISSN 0066-4197, 1545-2948. doi: 10.1146/annurev.genet.35.102401.091334. URL http://www.annualreviews.org/doi/10.1146/annurev.genet.35.102401.091334.

K. Nasmyth and C. H. Haering. Cohesin: Its Roles and Mechanisms. *Annual Review of Genetics*, 43(1):525–558, Dec. 2009. ISSN 0066-4197, 1545-2948. doi: 10.1146/annurev-genet-102108-134233. URL http://www.annualreviews.org/doi/10.1146/annurev-genet-102108-134233.

G. Nir, I. Farabella, C. Pérez Estrada, C. G. Ebeling, B. J. Beliveau, H. M. Sasaki, S. D. Lee, S. C. Nguyen, R. B. McCole, S. Chattoraj, J. Erceg, J. AlHaj Abed, N. M. C. Martins, H. Q. Nguyen, M. A. Hannan, S. Russell, N. C. Durand, S. S. P. Rao, J. Y. Kishi, P. Soler-Vila, M. Di Pierro, J. N. Onuchic, S. P. Callahan, J. M. Schreiner, J. A. Stuckey, P. Yin, E. L. Aiden, M. A. Marti-Renom, and C.-t. Wu. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLOS Genetics*, 14(12):e1007872, Dec. 2018. ISSN 1553-7404. doi: 10.1371/journal.pgen.1007872. URL http://dx.plos.org/10.1371/journal.pgen.1007872.

K. Nishimura, T. Fukagawa, H. Takisawa, T. Kakimoto, and M. Kanemaki. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods*, 6(12):917–922, Dec. 2009. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.1401. URL http://www.nature.com/articles/nmeth.1401.

T. Nishiyama, R. Ladurner, J. Schmitz, E. Kreidl, A. Schleiffer, V. Bhaskara, M. Bando, K. Shirahige, A. A. Hyman, K. Mechtler, and J.-M. Peters. Sororin Mediates Sister Chromatid Cohesion by Antagonizing Wapl. *Cell*, 143(5):737–749, Nov. 2010. ISSN 00928674. doi: 10.1016/j.cell.2010.10.031. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867410012353.

T. Nishiyama, M. M. Sykora, P. J. Huis in 't Veld, K. Mechtler, and J.-M. Peters. Aurora B and Cdk1 mediate Wapl activation and release of acetylated cohesin from chromosomes by phosphorylating Sororin. *Proceedings of the National Academy of Sciences*, 110(33):13404–13409, Aug.

2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1305020110. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1305020110.

E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–385, May 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11049. URL http://www.nature.com/articles/nature11049.

E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, and B. G. Bruneau. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, 169(5):930–944.e22, May 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.05.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417305317.

D. Normanno, L. Boudarène, C. Dugast-Darzacq, J. Chen, C. Richter, F. Proux, O. Bénichou, R. Voituriez, X. Darzacq, and M. Dahan. Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nature Communications*, 6:7357, July 2015. URL https://doi.org/10.1038/ncomms8357.

J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, and L. A. Mirny. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, 115(29):E6697–E6706, July 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1717730115. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1717730115.

R. Ohlsson, R. Renkawitz, and V. Lobanenkov. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in Genetics*, 17(9):520–527, Sept. 2001. ISSN 01689525. doi: 10.1016/S0168-9525(01)02366-6. URL https://linkinghub.elsevier.com/retrieve/pii/S0168952501023666.

H. D. Ou, S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O'Shea. ChromEMT: Visualizing 3d chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349):eaag0025, July 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aag0025. URL http://www.sciencemag.org/lookup/doi/10.1126/science.aag0025.

E. Passarge. Emil Heitz and the Concept of Heterochromatin: Longitudinal Chromosome Differentiation was Recognized Fifty Years Ago. page 10.

A. Pauli, J. G. van Bemmel, R. A. Oliveira, T. Itoh, K. Shirahige, B. van Steensel, and K. Nasmyth. A Direct Role for Cohesin in Gene Regulation and Ecdysone Response in Drosophila Salivary Glands. *Current Biology*, 20(20):1787–1798, Oct. 2010. ISSN 09609822. doi: 10.1016/j.cub.2010.09.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0960982210010870.

N. J. Petela, T. G. Gligoris, J. Metson, B.-G. Lee, M. Voulgaris, B. Hu, S. Kikuchi, C. Chapard, W. Chen, E. Rajendra, M. Srinivisan, H. Yu, J. Löwe, and K. A. Nasmyth. Scc2 Is a Potent Activator of Cohesin's ATPase that Promotes Loading by Binding Scc1 without Pds5. *Molecular Cell*, 70(6):1134–1148.e7, June 2018. ISSN 10972765. doi: 10.1016/j.molcel.2018.05.022. URL https://linkinghub.elsevier.com/retrieve/pii/S1097276518303964.

J.-M. Peters, A. Tedeschi, and J. Schmitz. The cohesin complex and its roles in chromosome biology. *Genes & Development*, 22(22):3089–3114, Nov. 2008. ISSN 0890-9369. doi: 10.1101/gad.1724308. URL http://genesdev.cshlp.org/cgi/doi/10.1101/gad.1724308.

J. E. Phillips and V. G. Corces. CTCF: Master Weaver of the Genome. *Cell*, 137(7):1194–1211, June 2009. ISSN 00928674. doi: 10.1016/j.cell.2009.06.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867409006990.

H. Qian, M. Sheetz, and E. Elson. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophysical Journal*, 60 (4):910–921, Oct. 1991. ISSN 00063495. doi: 10.1016/S0006-3495(91)82125-7. URL https://linkinghub.elsevier.com/retrieve/pii/S0006349591821257.

S. Rankin. Sororin, the Cell Cycle and Sister Chromatid Cohesion. *Cell Cycle*, 4(8):1039–1042, Aug. 2005. ISSN 1538-4101, 1551-4005. doi: 10.4161/cc.4.8.1926. URL https://www.tandfonline.com/doi/full/10.4161/cc.4.8.1926.

S. Rao, M. Huntley, N. Durand, E. Stamenova, I. Bochkov, J. Robinson, A. Sanborn, I. Machol, A. Omer, E. Lander, and E. Aiden. A 3d Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 159(7):1665–1680, Dec. 2014. ISSN 00928674. doi: 10.1016/j.cell.2014.11.021. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867414014974.

S. S. Rao, S.-C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K.-R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, and E. L. Aiden. Cohesin Loss Eliminates All Loop Domains. *Cell*, 171(2):305–320.e24, Oct. 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.09.026. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417311200.

J. Rhodes, D. Mazza, K. Nasmyth, and S. Uphoff. Scc2/Nipbl hops between chromosomal cohesin rings after loading. *Genes and Chromosomes*, page 20.

A. D. Riggs. DNA Methylation and Late Replication Probably Aid Cell Memory, and Type 1 DNA Reeling Could Aid Chromosome Folding and Enhancer Function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 326(1235):285–297, Jan. 1990. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.1990.0012. URL http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.1990.0012.

A. Rosa and R. Everaers. Structure and Dynamics of Interphase Chromosomes. *PLoS Computational Biology*, 4(8):e1000153, Aug. 2008. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000153. URL http://dx.plos.org/10.1371/journal.pcbi.1000153.

M. J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–796, Oct. 2006. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth929. URL http://www.nature.com/articles/nmeth929.

A. L. Sanborn, S. S. P. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, and E. L. Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, 112(47):E6456–E6465, Nov. 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1518552112. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1518552112.

M. Saxton. Single-particle tracking: the distribution of diffusion coefficients. *Biophysical Journal*, 72(4):1744–1753, Apr. 1997. ISSN 00063495. doi: 10.1016/S0006-3495(97)78820-9. URL https://linkinghub.elsevier.com/retrieve/pii/S0006349597788209.

W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, C. H. Haering, L. Mirny, and F. Spitz. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551(7678):51–56, Nov. 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature24281. URL http://www.nature.com/articles/nature24281.

A. Sergé, N. Bertaux, H. Rigneault, and D. Marguet. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nature Methods*, 5:687, July 2008. URL https://doi.org/10.1038/nmeth.1233.

T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*, 148(3):458–472, Feb. 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.01.010. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867412000165.

E. Smith, B. Lajoie, G. Jain, and J. Dekker. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *The American Journal of Human Genetics*, 98(1):185–201, Jan. 2016. ISSN 00029297. doi: 10.1016/j.ajhg.2015.12.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0002929715004978.

I. Solovei, K. Thanisch, and Y. Feodorova. How to rule the nucleus: divide et impera. *Current Opinion in Cell Biology*, 40:47–59, June 2016. ISSN 09550674. doi: 10.1016/j.ceb.2016.02.014. URL https://linkinghub.elsevier.com/retrieve/pii/S0955067416300217.

J. Stadler and H. Richly. Regulation of DNA Repair Mechanisms: How the Chromatin Environment Regulates the DNA Damage Response. *International Journal of Molecular Sciences*, 18(8):1715, Aug. 2017. ISSN 1422-0067. doi: 10.3390/ijms18081715. URL http://www.mdpi.com/1422-0067/18/8/1715.

J. Stigler, G. Çamdere, D. E. Koshland, and E. C. Greene. Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Reports*, 15(5):988–998, May 2016. ISSN 22111247. doi: 10.1016/j.celrep.2016.04.003. URL https://linkinghub.elsevier.com/retrieve/pii/S2211124716304028.

F. W. Studier and P. K. Bandyopadhyay. Model for how type I restriction enzymes select cleavage sites in DNA. *Proceedings of the National Academy of Sciences*, 85(13):4677–4681, July 1988. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.85.13.4677. URL http://www.pnas.org/cgi/doi/10.1073/pnas.85.13.4677.

Q. Szabo, D. Jost, J.-M. Chang, D. I. Cattoni, G. L. Papadopoulos, B. Bonev, T. Sexton, J. Gurgo, C. Jacquier, M. Nollmann, F. Bantignies, and G. Cavalli. TADs are 3d structural units of higher-order chromosome organization in *Drosophila*. *Science Advances*, 4(2):eaar8082, Feb. 2018. ISSN 2375-2548. doi: 10.1126/sciadv.aar8082. URL http://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aar8082.

Q. Szabo, F. Bantignies, and G. Cavalli. Principles of genome folding into topologically associating domains. *Science Advances*, 5(4):eaaw1668, Apr. 2019. ISSN 2375-2548. doi: 10.1126/sciadv.aaw1668. URL http://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aaw1668.

A. Tedeschi, G. Wutz, S. Huet, M. Jaritz, A. Wuensche, E. Schirghuber, I. F. Davidson, W. Tang, D. A. Cisneros, V. Bhaskara, T. Nishiyama, A. Vaziri, A. Wutz, J. Ellenberg, and J.-M. Peters. Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature*, 501(7468):564–568, Sept. 2013. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature12471. URL http://www.nature.com/articles/nature12471.

R. E. Thompson, D. R. Larson, and W. W. Webb. Precise Nanometer Localization Analysis for Individual Fluorescent Probes. *Biophysical Journal*, 82(5):2775–2783, May 2002. ISSN 00063495. doi: 10.1016/S0006-3495(02)75618-X. URL https://linkinghub.elsevier.com/retrieve/pii/S000634950275618X.

M. Tokunaga, N. Imamoto, and K. Sakata-Sogawa. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature Methods*, 5(2):159–161, Feb. 2008. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth1171. URL http://www.nature.com/articles/nmeth1171.

M. Tsochatzidou, M. Malliarou, N. Papanikolaou, J. Roca, and C. Nikolaou. Genome urbanization: clusters of topologically co-regulated genes delineate functional compartments in the genome of Saccharomyces cerevisiae. *Nucleic Acids Research*, 45(10):5818–5828, June 2017. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkx198. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkx198.

A.-L. Valton and J. Dekker. TAD disruption as oncogenic driver. *Current Opinion in Genetics & Development*, 36:34–40, Feb. 2016. ISSN 0959437X. doi: 10.1016/j.gde.2016.03.008. URL https://linkinghub.elsevier.com/retrieve/pii/S0959437X16300168.

B. van Steensel and A. S. Belmont. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*, 169(5):780–791, May 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.04.022. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417304737.

L. Vian, A. Pekowska, S. S. Rao, K.-R. Kieffer-Kwon, S. Jung, L. Baranello, S.-C. Huang, L. El Khattabi, M. Dose, N. Pruett, A. L. Sanborn, A. Canela, Y. Maman, A. Oksanen, W. Resch, X. Li, B. Lee, A. L. Kovalchuk, Z. Tang, S. Nelson, M. Di Pierro, R. R. Cheng, I. Machol, B. G. St Hilaire, N. C. Durand, M. S. Shamim, E. K. Stamenova, J. N. Onuchic, Y. Ruan, A. Nussenzweig, D. Levens, E. L. Aiden, and R. Casellas. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*, 173(5):1165–1178.e20, May 2018. ISSN 00928674. doi: 10.1016/j.cell.2018.03.072. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867418304045.

Y. Wang, F. Wang, R. Wang, P. Zhao, and Q. Xia. 2a self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. *Scientific Reports*, 5(1):16273, Dec. 2015. ISSN 2045-2322. doi: 10.1038/srep16273. URL http://www.nature.com/articles/srep16273.

J. N. Wells, T. G. Gligoris, K. A. Nasmyth, and J. A. Marsh. Evolution of condensin and cohesin complexes driven by replacement of Kite by Hawk proteins. *Current Biology*, 27(1):R17–R18, Jan. 2017. ISSN 09609822. doi: 10.1016/j.cub.2016.11.050. URL https://linkinghub.elsevier.com/retrieve/pii/S096098221631404X.

K. S. Wendt and J.-M. Peters. How cohesin and CTCF cooperate in regulating gene expression. *Chromosome Research*, 17(2):201–214, Feb. 2009. ISSN 0967-3849, 1573-6849. doi: 10.1007/s10577-008-9017-7. URL http://link.springer.com/10.1007/s10577-008-9017-7.

K. S. Wendt, K. Yoshida, T. Itoh, M. Bando, B. Koch, E. Schirghuber, S. Tsutsumi, G. Nagae, K. Ishihara, T. Mishiro, K. Yahata, F. Imamoto, H. Aburatani, M. Nakao, N. Imamoto, K. Maeshima, K. Shirahige, and J.-M. Peters. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, 451(7180):796–801, Feb. 2008. ISSN 0028-0836,

1476-4687. doi: 10.1038/nature06634. URL http://www.nature.com/articles/nature06634.

G. Wutz, C. Várnai, K. Nagasaka, D. A. Cisneros, R. R. Stocsits, W. Tang, S. Schoenfelder, G. Jessberger, M. Muhar, M. J. Hossain, N. Walther, B. Koch, M. Kueblbeck, J. Ellenberg, J. Zuber, P. Fraser, and J. Peters. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *The EMBO Journal*, 36(24):3573–3599, Dec. 2017. ISSN 0261-4189, 1460-2075, 0261-4189, 1460-2075. doi: 10.15252/embj.201798004. URL http://emboj.embopress.org/lookup/doi/10.15252/embj.201798004.

Y. Zhan, L. Mariani, I. Barozzi, E. G. Schulz, N. Blüthgen, M. Stadler, G. Tiana, and L. Giorgetti. Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Research*, 27 (3):479–490, Mar. 2017. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.212803.116. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.212803.116.

A. Zidovska, D. A. Weitz, and T. J. Mitchison. Micron-scale coherence in interphase chromatin dynamics. *Proceedings of the National Academy of Sciences*, 110(39):15555–15560, Sept. 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1220313110. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1220313110.

J. Zuin, J. R. Dixon, M. I. J. A. van der Reijden, Z. Ye, P. Kolovos, R. W. W. Brouwer, M. P. C. van de Corput, H. J. G. van de Werken, T. A. Knoch, W. F. J. van IJcken, F. G. Grosveld, B. Ren, and K. S. Wendt. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences*, 111(3):996–1001, Jan. 2014. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1317788111. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1317788111.

## RÉSUMÉ

La structure de la chromatine joue un rôle crucial dans la régulation de plusieurs fonctions cellulaires chez les cellules de mammifères. Perturber l'organisation spatiale de la chromatine peut avoir des conséquences dramatiques sur la vie d'une cellule et peut amener à des pathologies graves chez les organismes. Deux facteurs nucléaires, CTCF et Cohesine, sont parmi les principaux acteurs de la régulation et du maintien de l'architecture de l'ADN. Des avancements importants ont révélé la complexité des mécanismes qui régulent l'organisation de la chromatine, mais le domaine manque encore d'une description dynamique à l'échelle de la cellule et de la molécule unique. Cette étude est centrée sur la description de la dynamique de CTCF et Cohesin, réalisée avec des méthodes de suivi de la molecule unique dans des cellules souches embryonnaires vivantes de souris. L'intéraction entre ces deux facteurs a été étudiée à travers la caractérisation de la dynamique de Cohesin en absence de CTCF et dans le contexte d'autres altérations biologiques. Les résultats obtenus montrent que CTCF n'intervient pas dans la régulation de la dynamique de liaison de Cohesine à l'ADN. D'autre part, on observe une faible mais systématique augmentation du coefficient de diffusion de la Cohesine en absence de CTCF. Ensemble, ces résultats supportent l'hypothèse selon laquelle CTCF bloque les Cohesines en train d'extruder des boucles de chromatine.

## MOTS CLÉS

Chromatin, dynamique, CTCF, Cohesin, structure et fonction, suivi de molecule unique, microscopie optique, TAD, domaine d'association topologique.

## ABSTRACT

Chromatin structure and cellular function are tightly linked in the nucleus of mammalian cells. Disruption of chromatin spatial organisation dramatically affects the life of a cell and eventually leads to severe pathologies in entire organisms. Two nuclear factors, CTCF and Cohesin, have been found to play a crucial role in the regulation and maintenance of DNA architecture. Huge advancements have been made in the understanding of the mechanisms behind chromatin arrangement but the field is still lacking a dynamic picture at the single cell and single molecule level. This study provides insight into the dynamics of CTCF and Cohesin through single particle tracking of CTCF and Cohesin dynamics achieved with single molecule tracking in living mouse embryonic stem cells. The interplay between these two factors was studied by looking at Cohesin's behaviour in the absence of CTCF and in the context of other biological alterations. Our results show that CTCF does not play a role in the regulation of Cohesin's binding kinetics. On the other hand we observe a mild but systematic increase of Cohesin's diffusion coefficient upon CTCF depletion. Together, these findings support the hypothesis for which CTCF acts a stop signal for loop-extruding Cohesins.

## KEYWORDS

Chromatin, dynamics, CTCF, Cohesin, structure and function, single molecule tracking, optical microscopy, TAD.