Tools to analyze single-particle tracking data in mammalian cells

Maxime Woringer

Sorbonne Université

Ecole doctorale 515 : Complexité du Vivant

Imaging and Modeling Unit (Institut Pasteur, Paris)

Tjian-Darzacq lab (UC Berkeley)

Tools to analyze single-particle tracking data in mammalian cells

Par Maxime Woringer

Thèse de doctorat de Biologie

Dirigée par Christophe Zimmer et Xavier Darzacq

Présentée et soutenue publiquement le 17 juin 2019

Devant un jury composé de :

Angela Taddei, directrice de Recherches, en tant que présidente du jury,

Christophe Zimmer, directeur de Recherches, en tant que directeur de thèse,

Xavier Darzacq, directeur de recherches, en tant que co-directeur de thèse,

Davide Mazza, directeur de recherches, en tant que rapporteur,

Marcelo Nollmann, directeur de recherches, en tant que rapporteur,

Katharina Proksch, chargée de recherches, en tant qu'examinatrice,

Antoine Coulon, chargé de recherches, en tant qu'examinateur,

Denis Grebenkov, chargé de recherches, en tant qu'examinateur.

Tools to analyze single-particle tracking data in mammalian cells

Maxime Woringer

May 27, 2019

Contents

Acknowledgements

I would like to warmly thank my PhD supervisors, Christophe, Xavier for their continuous support and mentorship, and all the members of both the IMOD lab and the Tjian-Darzacq lab, including former lab members: Sheila, Xian, Wei, Frank and Astou. I want to have a special thank to all my colleagues that accepted to collaborate with me on different projects: Anders, Jyotsana, Christian, Mickaël, Benoît, Mustafa, Ignacio, Cyril, Hugues, Davide and also Tatiana, Inma, Philippe, Zhijie, Alan, Kartoosh, Korai, Jean-Yves Tinevez, Jean-Yves Coppée and Odile. I also had the chance to initiate projects or have fascinating discussions with the labs of Pablo Navarro, Anne Dejean, Luke Lavis, Caterina Strambio, Susan Marqusee, Maxime Dahan and Antoine Coulon. I was very lucky to work with two talented undergraduate students, Josh and Stephan. I received a lot of help from the lab staff and administrative staff, including from Mallory, Janeen, Marie, Marie-Anne Lin, Gina and Ana. Finally, I am extremely grateful to everyone who provided moral and emotional support during difficult times, including Elena, Wulan and Florian.

Introduction

A key question in biology is to understand the relationship between structure and function. Although structures are being mapped with increasing precision using a whole range of techniques such as Cryo-EM or super-resolution microscopy, this link remains imperfect because we do not understand how dynamic features such as the motion of proteins are involved in mediating the structure-function relationship.

Proteins can exist in many states, that cannot be resolved using bulk assays and single-molecule techniques such as single-particle tracking (SPT) have been developed to track the motion of proteins in live cells. SPT has experienced a growing attention since its inception in live cells in the early 2000s. Despite two decades of development, however, it is still unclear how SPT data should be analyzed.

In this dissertation, we present several projects developed during my PhD:

1. The development of SPT analysis tools, that can be used to analyze real-life SPT data. We first implemented a propagator-based method, called Spot-On, and made it widely available through a web-interface.

2. The implementation of a simulation tool to benchmark Spot-On, that incorporates realistic biases and can simulate the motion of diffusing particles under various types of motion.

3. The evaluation of Spot-On in an *in vitro* SPT setting, in an application related to catalysis-enhanced diffusion.

4. The preliminary analysis of the diffusion behaviour of the transcription factor c-Myc, with a focus on its biophysical properties.

Part I

Background

Chapter 1

Short introduction: why do we care about protein diffusion?

In this section we provide a quick and broad introduction/entry-point to motivate the rest of the dissertation. We explicitly refer the reader to more detailed sections or introduction when needed.

1.1 The regulation of biological processes

The main focus of this dissertation is to produce tools to dissect the regulation of biological processes. In very abstract terms, one can see a biological process as a chemical reaction that turns one or many reagents (for instance A and B) into one or many products (C), leading to the formal chemical reaction:

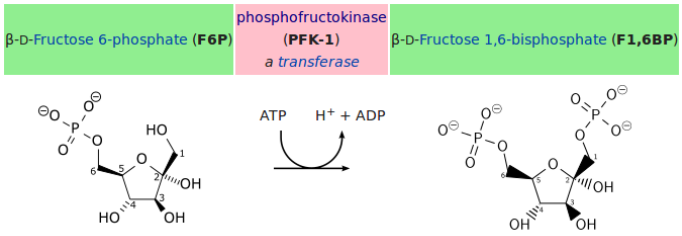
$$A + B \rightarrow C$$

Chemical reactions are usually associated with a kinetic rate, or kinetic constant k. To say that a biological process is regulated means that this kinetic constant is not a real constant, but a function of the concentration of external factors: $k = k([\text{external factors}])$. Canonical examples of regulated processes in biology include the regulation of the first step of glycolysis, during which the enzyme Phospho-Fructo-Kinase (PFK1) catalyzes the phosphorylation of its substrate, the fructose-6-phosphate in fructose-1,6-bisphosphate (Figure 1.1a). This step is regulated by the concentration of many factors, termed allosteric regulators.

Another, simplified example is the regulation of the Lac operon (Pardee, Jacob, and Monod 1959): the formal reaction that turns a DNA segment, the bacterial RNA polymerase and many other cellular factors into a mRNA product (the expression of the LacZ gene; Figure 1.1b). The expression of this gene is known to be highly regulated, and one factor that regulates the rate of mRNA production is the concentration of lactose.

A third example, closer to the topic of this dissertation, is the regulation of transcription in mammalian cells. (Figure 1.1c). In mammalian cells, the production of a mRNA depends on the step-wise assembly of a huge multi-protein complex termed the transcription pre-initiation complex (PIC; presented in section I.3). Each step of the assembly is regulated by the presence/absence or the concentration of molecules named transcription factors (TFs). As such, transcription regulation is another example of regulated biological process.

a. k([allosteric regulators])



b. k([lactose])

$$\text{RNAP} + \text{DNA} + \ldots \longrightarrow \beta\text{−galactosidase (LacZ)}$$

c. k_{mRNA}([transcription factors])

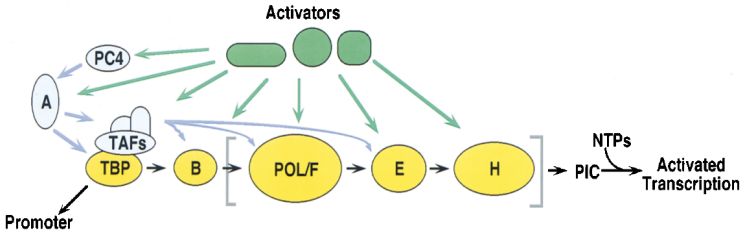


Figure 1.1: **(a)** The first step of glycolysis is regulated by the concentration of many allosteric modulators. **(b)** In bacterial systems, the expression of the LacZ gene in the Lac operon is regulated by the concentration of lactose in the medium. **(c)** In mammalian systems, the expression of a messenger RNA (mRNA) is regulated by the presence/absence of many TFs that can affect the step-wise assembly of the PIC at many critical steps (source: Nikolov and S. K. Burley 1997 and Wikipedia).

1.2 Why diffusion matters

According to traditional kinetics, once the partners (concentration of reagents and products) involved in a bio-molecular reaction and the function that links the concentration of regulatory molecules to the reaction rate are known, the concentration and time evolution of the reagents and products is known.

However, we know since 1906, thanks to the Polish chemist Marian von Smoluchowski, that in a dilute medium, the reaction rate k (more specifically, the association rate of the reactants or diffusion limited on-rate) of a bi-molecular reaction can be decomposed into:

$$k = 4\pi DA,$$

where D is the (sum of the) diffusion coefficients of the reacting species, and A is a factor that has the unit of space, and that incorporates information about the relative conformation of the two reacting species, and can be used to reflect the chemistry of the molecules.

This finding is absolutely critical. It means that to determine the reaction rate of *any* bio-molecule, one cannot only have some information about the partners involved in the reaction, but also about the space in which these particles diffuse. If for some reason they diffuse faster, then the reaction rate is increased. If they diffuse slower, the reaction rate is decreased. Thus, transcription can be directly regulated by tuning the diffusion coefficient of one molecular species. Section 3.3 details how the spatial organization of the nucleus can affect transcription.

Furthermore, one can compare two situations: a first one in which a protein diffuses freely into a confinement volume (such as the nucleus; illustrated in Figure 1.2a-left), and one in which the protein diffuses in the same volume, but the volume is randomly and uniformly filled with obstacles. In real life, these obstacles can be the cytoskeleton, DNA and chromatin, nucleoli, phase-separated compartments, etc. In such a situation, one can mathematically and numerically show that a key parameter to characterize diffusion, the *mean first passage time* (MFPT) shows qualitative differences:

- In the case of the empty volume, the protein has a constant MFPT across space (Figure 1.2b-left): regions of space that are close and far from the starting point are equally likely to be visited, and in this situation, it does not really matter whether the diffusing protein and the target are far apart: distance does not matter. This type of exploration is termed *non-compact*, since all sites are visited with equal probability.

- In the case of the volume filled with obstacles, *distance matters* (Figure 1.2b-right): the MFPT is much lower for regions close in space than in regions far in space. This type of exploration is termed *non-compact*.

If diffusion of TFs in cell were *non-compact*, then spatial proximity relationships in the nucleus would not really matter. It would be very difficult to explain why co-regulated genes tend to co-localize genomically, why the genome is organized in topologically-associated domains (TADs) and why enhancer-promoters come into contact. As such, knowing about the structure of the space, and how it is filled with obstacles actually provide critical information about how proteins diffuse in cells, and how this might affect transcription.

To sum up, one can see this *mean first passage time* using the following analogy: a chocolate cake (the target) is hidden in a complex space (a medium filled with obstacles), such as the house in Figure 1.2c, a complex medium with lots of different floors, different stairs and filled with people with funny hats. When kids (the diffusing proteins) search for the cake (pictured in Figure 1.2d), it takes them a long time to find it, because the house is intricate and complicated. However, when the kids encounter the cake, the reaction takes place instantly and the cake "magically disappear".

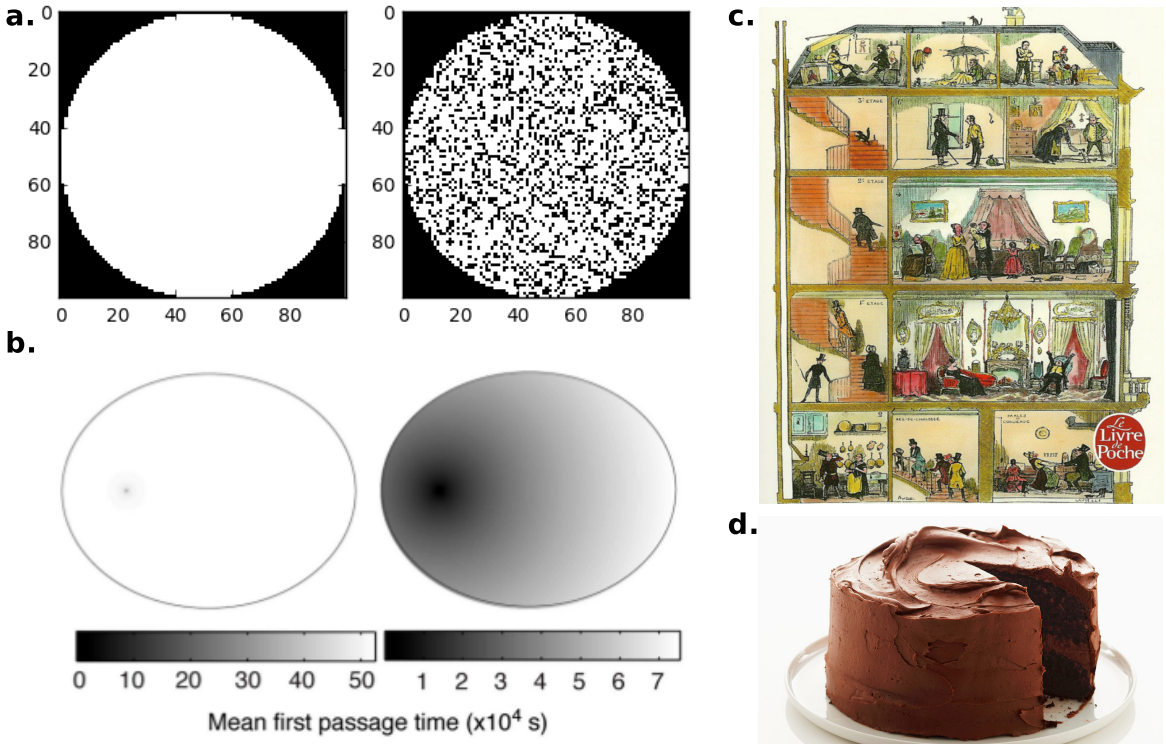


Figure 1.2: **(a).** Two diffusion volumes *(left)* an empty container, in which molecules can diffuse freely and *(right)* a container full of –randomly placed– obstacles that hinder diffusion. **(b).** Representation of the mean first passage time for the two geometries presented, (from Izeddin, Récamier, et al. 2014). **(c).** A metaphor for a complex geometry: a house with a lot of floors, stairs, rooms and people with funny hats, (from La Vie Mode d'Emploi, Georges Perec), **(d).** An instantly reacting species: the chocolate cake is instantly catalyzed into stomach-ache when the children-macro-molecular complex encounters it.

1.3 The nucleus as a complex space

In biology, the nucleus is a good candidate for such an intricate space, in which the spatial organization of chromatin and the nucleus can significantly affect diffusion. Section I.2 highlights some features of nuclear organization in more details. Briefly, each human cell contains approximately two meters of DNA that are folded in a 10 µm nucleus (section 2.2). This folding takes place at many levels: nucleosomes (section 2.3), TADs (section 2.4) and compartments (section 2.5). In addition to chromatin, mammalian nuclei contain a hierarchy of macro-molecular structures (pictured in Figure 1.3), many of them are now termed membrane-less organelles. Membrane-less organelles include, among many others, nuclear speckles, nucleoli and PML bodies. These organelles bear various chemical properties and can differently affect diffusion.

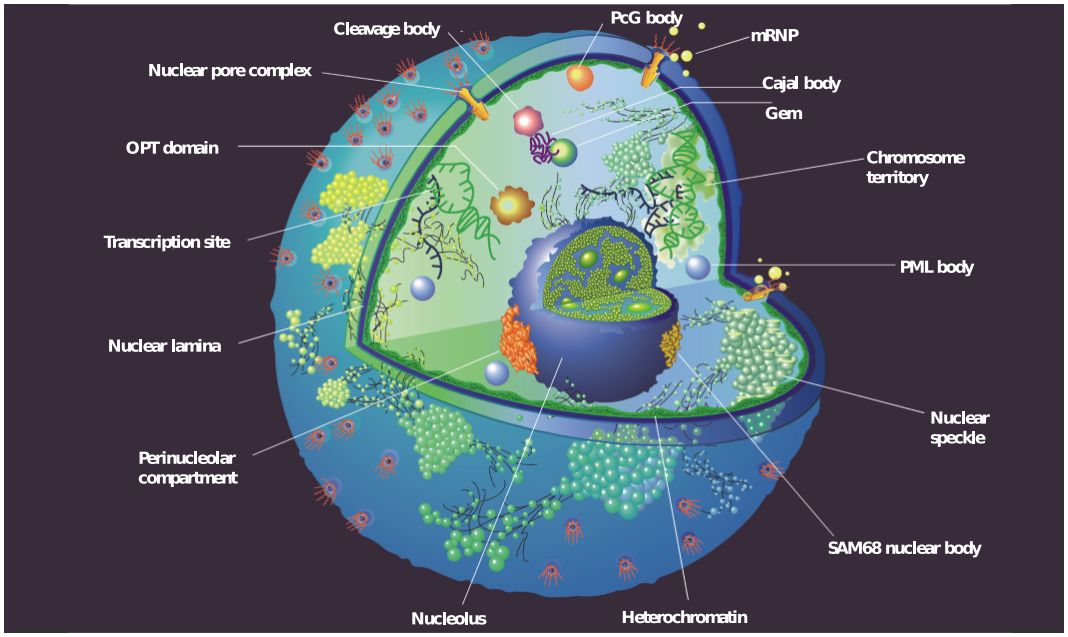


Figure 1.3: An overview of frequent membrane-less organelles and phase-separated bodies in a mammalian nucleus (source: David L Spector and Gasser 2003).

1.4 Single-particle tracking to study nuclear diffusion

In such a complex nucleus, there is a need to characterize how proteins diffuse. This will bring valuable insights on how nuclear organization an protein dynamics intertwine. Many techniques have been used to study diffusion of proteins in cells, and some of them are reviewed in section I.4. Among them, single-particle tracking (SPT) has emerged as a key technique, since it provides precise information about single molecule diffusion in a live cell context.

In a conventional SPT experiment (reviewed in more details in section 4.1), the protein of interest is labeled, for instance using a HaloTag fusion protein conjugated to a fluorescent dye (Figure 1.4a). Then using a microscope with single-molecule sensitivity, high-speed cameras and an ad-hoc illumination scheme (Figure 1.4b), the fluorescence emitted by the dye is recorded over the lifetime of the fluorescent molecule. If the labeling has been performed in a sufficiently low density (for instance if photo-activation has been used, or if the fluorescent dye is the limiting reagent in the conjugation reaction), single molecules appear as isolated spots that can be detected and tracked over time (Figure 1.4c), allowing to extract the coordinates of the spots, and to reconstruct trajectories (Figure 1.4d). The trajectories can then be analyzed using statistical methods to extract meaningful information about diffusion (Figure 1.4e).

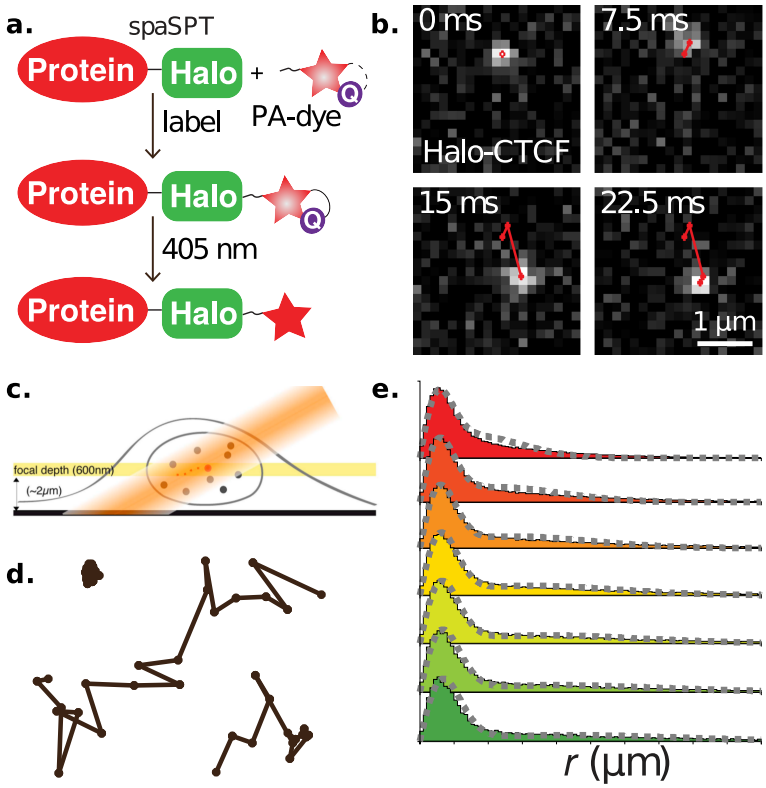


Figure 1.4: **A procedure to perform single-particle tracking.** **(a)** The protein of interest is labeled for instance using a HaloTag-fusion protein and a photo-activatable organic dye (PA-dye). Under UV illumination, this dye can be converted into a fluorescent state. **(b)** Single molecules can then be tracked over time under a fluorescence microscope, using a detection and fitting algorithm followed by a tracking method. **(c)** To see single molecules, specific illumination schemes usually need to be employed in order to reduce out-of-focus illumination. One of them, depicted in the picture, is HiLo illumination (Tokunaga, Imamoto, and Sakata-Sogawa 2008). **(d)** Trajectories can then be extracted fro the movies and **(e)** statistical analysis, here modeling with Spot-On can be performed (source: Anders S. Hansen, Woringer, et al. 2018).

The analysis of SPT is the subject of intense research, and we present an example of analysis technique, Spot-On, in section 1.3. An overview of existing analysis techniques is presented in section 4.4.

Chapter 2

General structure and function of the nucleus

Understanding how transcription factors (TFs) regulate mammalian gene expression in space and time is a central topic in biology. To activate a gene, a TF has first to diffuse in the available space of the nucleus until it reaches a target DNA sequence or protein (target site). This eventually results in the recruitment of the whole transcriptional machinery.

All these processes take place in the mammalian nucleoplasm, a highly organized and dynamic environment, in which some complexes transiently assemble and break apart, whereas others appear more stable. This diversity of dynamic behaviors arises from the number of bio-molecules that make up the nucleoplasm and their pairwise interactions. Indeed, interactions energies that span several orders of magnitude, from covalent bounds to transient and dynamic interactions can shape nuclear landscapes. Thus, the nuclear environment determines how frequently and how fast a TF contacts its target site, and it indirectly regulates gene expression.

Among the nuclear constituents is the genome. Mammalian genomes in particular are under considerable constraints. The DNA double helix has a diameter of ~ 2 nm, and in human cells, its total length if stretched out would be ~ 2 m, but it is folded inside a nucleus ~ 10 µm in diameter, roughly five orders of magnitude smaller. Despite this tight folding, the genome needs to remain accessible to key biological processes, including DNA replication and gene expression. How genome architecture and biological processes intertwine has puzzled generations of biologists, and a more comprehensive picture of the key determinants of these interactions is just starting to be unraveled. Over the last decade, the field has undergone a dramatic acceleration thanks to the development of powerful sequencing-based assays and microscopy techniques, which have revealed previously unknown levels of chromatin organization. Together with polymer-based modeling, these data have helped to uncover some of the fundamental mechanisms that shape chromatin organization.

In this section, we review some general knowledge about the structure and function of the nucleus, with a focus on chromatin organization and membrane-less organelles. We try to simultaneously highlight the biophysical mechanisms leading to the current organization of the nucleus. The goal of this section is to set the stage where diffusion occurs. Concretely, section 2.1 draws a broad overview of our current understanding of "what is in the nucleus". The next sections progressively details how chromatin is organized in the nucleus.

Most of the content is extracted from three reviews that we wrote with Jyotsana, Ignacio and Xavier:

- "Geometry of the nucleus: a perspective on gene expression regulation", published in Current Opinion in Chemical Biology (Woringer, Darzacq, and Izeddin 2014),
- "Protein motion in the nucleus: from anomalous diffusion to weak interactions", published in Biochemical Society Transactions (Woringer and Darzacq 2018),

- "How the Genome Folds: The Biophysics of Four-Dimensional Chromatin Organization", published in Annual Reviews of Biophysics (Parmar, Woringer, and Zimmer 2019).

We apologize to our colleagues whose work could not be cited due to limited space. We thank Elena Rensen and Anders S. Hansen for critical feedback on the manuscript and Vincent Récamier, Raphaël Voituriez, Leonid Mirny, Yitzhak Rabin, Lana Bosanac and Benjamin Guglielmi.

2.1 What is in a nucleus?

Mammalian gene expression and its regulation take place in the nucleus, a highly complex and sub-compartmented organelle. Interactions strengths between nuclear constituents span several orders of magnitude, from covalent bounds to “strong” non-covalent interactions. These interactions lead to the formation of macro-molecular structures, either stable (Figure 2.1-left; for instance double-stranded DNA or biochemically purifiable macro-molecular complexes such as the ones involved in gene expression) or transient but specific, leading to preferential associations of classes of proteins (Figure 2.1-right). This section briefly highlight some of these constituents.

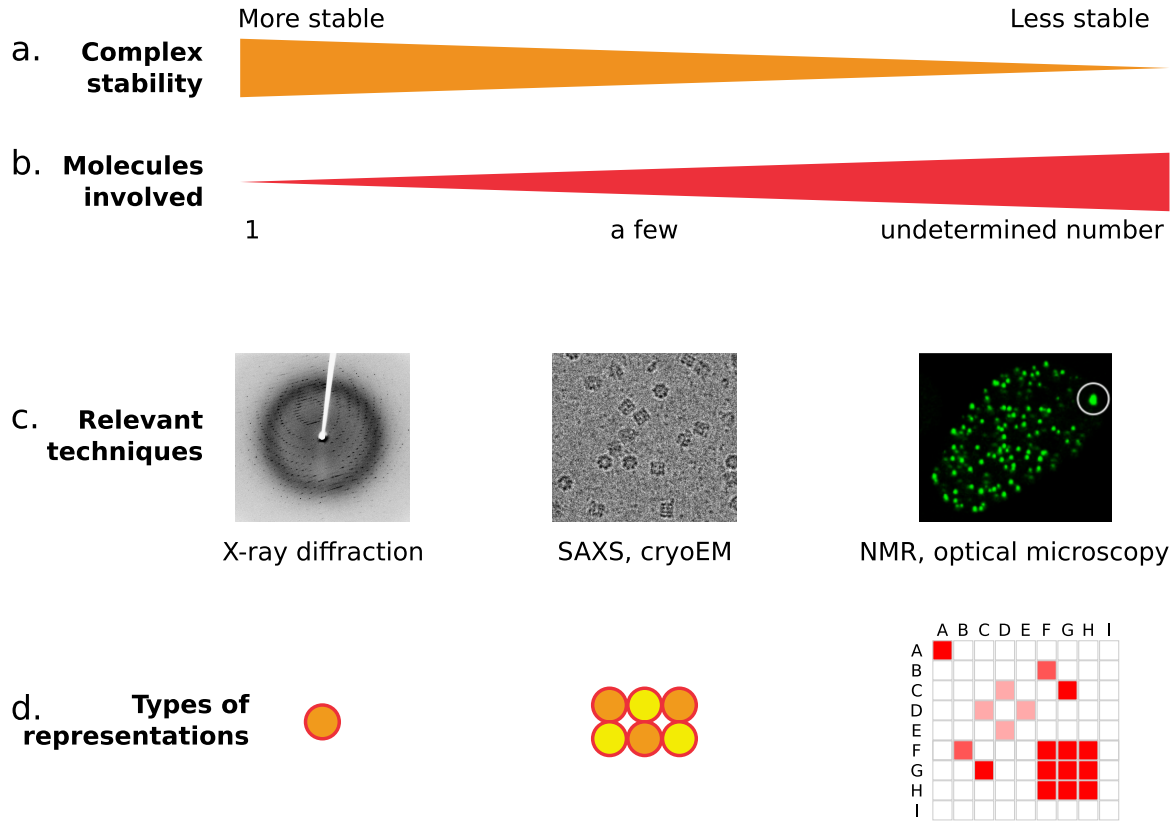


Figure 2.1: Biological interactions cover a wide spectrum in terms of complex stability **(a)** and number of molecules involved **(b)**. This spectrum spans from stable protein complexes that can be purified and further imaged by techniques such as X-ray diffraction *(left)* to very labile, transient interactions that can involve thousands of proteins *in vivo*, whereas none of the interactions can be captured by traditional biochemistry *(right)*. As one goes from one end of the spectrum to the other end, distinct sets of techniques **(c)** and types of representations **(d)** are needed. For instance, as the valency of interactions increases from a few strongly interacting partners to many weakly interacting partners, new graphical representations are needed, since traditional schematics representing macro-molecular complexes whose stoichiometry is known as the juxtaposition of monomers (*center*) become difficult to read when depicting one protein weakly interacting with dozen of partners (*right*). In that case, matrices of pairwise interactions between proteins A-I might be more relevant. SAXS: Small-Angle X-Rays Scattering, NMR: Nuclear Magnetic Resonance. Photographs are courtesy of Jeff Dahl, Vossman and Shasha Chong, respectively from left to right.

2.1.1 Content of a mammalian nucleus

It is surprisingly hard to find information on the macro-molecular content of the nucleus of a mammalian cell, and most of the information that I managed to find are highly approximate. Nonetheless, RNA, DNA, proteins are abundant in the nuclear space, and also some other elements such as ATP or aminoacids.

1. **Available space in the nucleus**. High molecular weight components in the nucleus, such as prominently but not exclusively chromatin, effectively reduce the accessible volume in which TFs are free to diffuse, potentially regulating the process of gene expression. A "rule of thumb" for the volume of a DNA is 1 nm^3/bp, according to `Bionumbers` (accession number 103778). Thus, neglecting adsorbed water, the volume of human DNA is $\sim$ 2x3x10^9 = 6x10^9 nm^3. Similarly, the exclusion volume of nucleosomes can be computed (crystal structure of the human nucleosome core, `10.2210/pdb2cv5/pdb`, and Bionumbers, accession numbers: 102977 and 102987), leading to an estimated volume of chromatin of $\sim$ 25 μm^3, which is a fraction of 12% of the volume of

a human nucleus ($\sim$ 6 µm diameter). Other estimates (10% in Rippe 2007, 20–50% in Bancaud et al. 2009) give similar order of magnitude. In a simple model of first order reaction, such exclusion volume would at most change by a mere factor of two the rate of homogeneous biochemical reactions. We must thus take into consideration other characteristics such as the complex geometry of nuclear organization or the heterogeneity of local molecular concentration. The former, as discussed below, renders the calculations of exclusion volume invalid; regarding the latter, many nuclear components do not show a homogeneous spatial distribution in the nucleus (Dowen et al. 2013), and it has been shown that the local concentration of Pol II is regulated, giving rise to significant differences at the local level throughout the nucleoplasm (Cisse et al. 2013).

2. **Heterogeneities in the nucleus.** At a finer scale, many of the constituents of the nucleus are heterogeneous, giving rise to changes in concentration. The main heterogeneities studied consider the alternance between DNA-rich and DNA-poor regions, investigated by super-resolution microscopy (Szczurek et al. 2014; Récamier et al. 2014) or Cryo-EM (Ou et al. 2017). Furthermore, it was proposed (T. Cremer, M. Cremer, et al. 2015) and recently evidenced that DNA-rich and RNA-rich regions are spatially segregated, based on SIM and STED imaging (Miron et al. 2019; Hilbert et al. 2017). Finally, provided the potential of multivalent RNA to facilitate phase-separation (see for instance Jain and Vale 2017 for a recent example), the interplay between DNA, RNA and proteins becomes of paramount importance, and several efforts have been made to map DNA-RNA interactions (see for instance the ChAR-seq and other techniques; Bell, Jukam, Teran, Risca, O. K. Smith, Johnson, J. Skotheim, et al. 2017; Morf et al. 2017).

2.1.2 Weak interactions in the nucleus

Unlike inert tracers whose diffusion is only determined by volume exclusion, proteins have both a relevant shape and charge pattern that determine their interaction landscape and thus their diffusive properties. These non-covalent interactions are obviously crucial to form biochemically stable complexes such as the transcription pre-initiation complex or the spliceosome (Figure 2.1-left), but also to form dynamic emergent structures of reduced dimensionality upon which TFs can transiently adsorb and diffuse (Figure 2.1-right). Under this model, proteins do not form stable complexes anymore, but rather have a high number of weakly-interacting partners. Indeed, simulation studies have shown that phase separation yielding structures of reduced dimensionality can occur under very minimal hypotheses, such as weak overall protein-protein attraction (Osmanović and Rabin 2016; Shagolsem and Rabin 2016).

In this context, the traditional representation of protein-protein interaction networks as graphs and arrows becomes less informative as the protein network gets fully connected, and can be replaced by representations such as pairwise interaction matrices (Figure 2.1d-right) (Bergeron-Sandoval, Safaee, and Michnick 2016).

Furthermore, the list of proteins exhibiting phase separation (that is, a state of matter in which part of the soluble protein fraction segregates into a liquid or liquid-like droplet) *in vitro* or *in vivo* is quickly growing, supporting the vision that the emergence of structures of reduced dimensionality is closer to a general organizing principle than an anecdotal biophysical phenomenon, and some of them were linked with transcriptional regulation (A. Tsai et al. 2017; Kwon et al. 2013; A. G. Larson et al. 2017; Sherry et al. 2017; S. Chong, Dugast-Darzacq, Z. Liu, Dong, G. Dailey, et al. 2017).

1. **Implications of phase separation.** First, transient structures of reduced dimensionality, including protein aggregates, phase separated domains or sub-nuclear compartments require at least one *multivalent partner*, that can nucleate the aggregation. As such, many abundant constituents of the mammalian nucleus have been shown to nucleate a structure of reduced dimensionality. These structures include the formation of nucleoli in which diffusion is highly constrained, phase-separation of heterochromatin protein 1 domains (HP1; A. G. Larson et al. 2017; Strom et al. 2017) and highly active chromatin domains, whose center has reduced accessibility to diffusing

proteins, restricting most diffusing proteins to the surface of the domain (S. Chong, Dugast-Darzacq, Z. Liu, Dong, G. Dailey, et al. 2017). The constituents include low complexity protein domains, that constitute the majority of the mammalian proteome (Shammas 2017; Kato et al. 2012), especially TFs (Jiangang Liu et al. 2006), repeated DNA (Nott et al. 2015) or RNA sequences (Jain and Vale 2017; P. Li et al. 2012; Molliex et al. 2015) or small amphiphilic molecules (Patel, Malinovska, et al. 2017; Patel, H. O. Lee, et al. 2015). Once nucleated, these structures can have a highly tortuous, potentially fractal architecture, and can serve as a scaffold that can weakly trap other proteins, leading to anomalous diffusion.

Second, the partners have to exhibit compatible interactions: it is chemically unlikely that both highly charged and hydrophobic proteins will coexist in the same structure without the help of additional compounds acting as counter-ions (Pak et al. 2016), setting the basis of a "grammar of interactions" (Gimona 2006), that is being progressively deciphered (Brady et al. 2017; Reichheld et al. 2017; Rahul K Das, Ruff, and Rohit V Pappu 2015; Sherry et al. 2017; Patel, H. O. Lee, et al. 2015; Quiroz and Chilkoti 2015).

Third, structures of reduced dimensionality emerging from weak interactions exhibit the following properties: (1) they usually exist as an extremely dynamic equilibrium rather than a stable structure (Strom et al. 2017; Wei et al. 2017; Molliex et al. 2015), and can thus be at the same time prevalent in the nucleus and hard to purify by traditional biochemistry that preferentially capture stable interactions. (2) Moreover, they emerge from a dynamic mesh of pairwise chemical interactions. They can show a high level of specificity, and several structures of reduced dimensionality can coexist in the same nucleus without intermixing (S. Chong, Dugast-Darzacq, Z. Liu, Dong, G. Dailey, et al. 2017; Shav-Tal et al. 2005; Pak et al. 2016; Kwon et al. 2013; Nott et al. 2015). Furthermore, the number and spatial relationships of such structures is only limited by the combinatorics of chemical interactions. (3) Finally, these structures can be regulated by the well-studied post-translational machinery of eukaryotic cells. For instance, phosphorylation of one of the proteins involved in such structure can trigger the timely disassembly of the whole structure and free all the factors interacting with it (Cisse et al. 2013; Kwon et al. 2013; J. T. Wang et al. 2014; Cho, Jayanth, Mullen, et al. 2016; A. G. Larson et al. 2017). All those factors will then exhibit a dramatically different dynamics and target search properties, potentially switching from a compact exploration mode to a non-compact one. As such, a specific (and potentially functional) group of factors can be regulated at once by modulation of the post-translational modifications of one scaffolding protein (P. Li et al. 2012; Nott et al. 2015).

2. **An example of weak interactions in the nucleus: the histone tails.** An recurrent question in nuclear biology is to what extent histone tails can behave as surfaces of reduced dimensionality.

   In most pictures we usually see, the histone tails are based on the nucleosome PDB structure 1AOI by Luger et al. 1997. It only has partial histone tails, and as such it cannot provide a full picture of the volume occupied by the tails. Conversely, (Arya and Schlick 2006) simulate histone tails in a nucleosome. They take full histone tails into account and it is clear that at a given instant the histone tails occupy a volume much lower than the core nucleosome (see the top panel of Figure 2.2). The simulation of histone tails, including histone tails modifications, is the subject of intense research (Grauffel, Stote, and Dejaegere 2010).

   On the other hand, the authors of (Arya and Schlick 2006) also simulate using a coarse grained model the "volume" occupied by histone tails (that is, the volume explored over time by the tail), and this volume more or less doubles the volume occupied by the nucleosome. It is unclear whether we could say that the capture radius of the nucleosome doubles (because the tail occupies only one conformation at a time). See the 4 bottom panels of Figure 2.2. The authors also quantify the extension of the tail, to be around 1-3 nm.

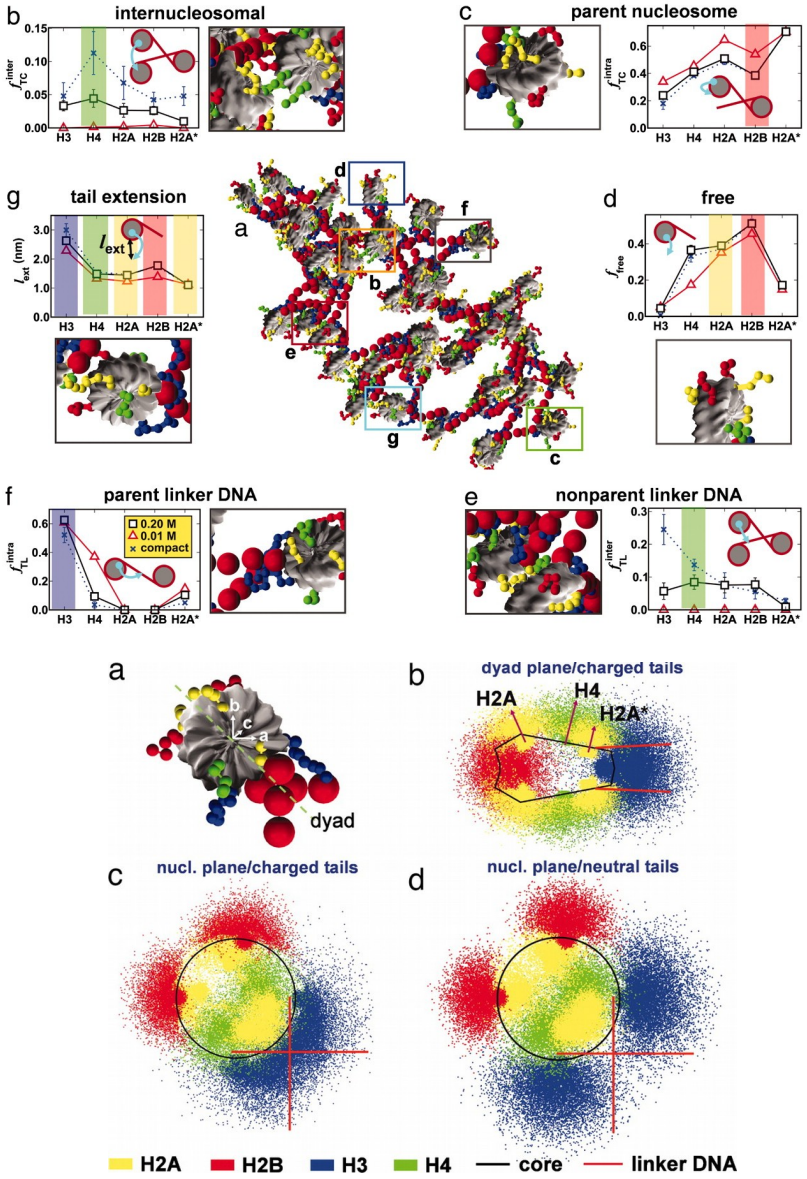


Figure 2.2: **Evaluation of the volume occupied by histone tails. top panels b-g** sample conformations and tail extension for various chromatin conformations. **bottom panels a-d** sample conformation and "volume" explored by histone tails over a long period of time (source: Arya and Schlick 2006).

3. **Conclusion.** The characterization of structures of reduced dimensionality emerging from weak interactions is still in its infancy, but appears more and more strongly as a clear organizing principle of mammalian nuclei. These structures create the matrix upon which fast-diffusing factors can specifically and transiently bind, diffuse and unbind, thus dynamically shaping the "diffusion landscape" of the whole transcriptional machinery. In the next sections, we review how chromatin is organized within the nucleus.

2.2 DNA and chromatin as polymers

The DNA double helix in a typical human chromosome consists of hundreds of millions of base pairs covalently chained together by sugar-phosphate backbones. Thus, DNA perfectly fits the definition of a polymer, as a molecule built from many similar elements (called monomers) bonded together. Polymers obey a wide range of properties that arise solely from the fact that they consist of many monomers and are largely independent of their precise chemical nature. This universality means that many concepts and results from polymer theory (Doi and Edwards 2007; Rubinstein and Colby 2003) can be applied to understand the structure and dynamics of DNA and chromatin in cells; we briefly recall some of them below.

A key property of polymers is semi-flexibility: On short length scales, a polymer behaves as a rigid rod, while on larger scales it can bend in arbitrary directions due to thermal agitation alone (Figure 2.3a). The length scale that separates rigid from flexible behavior is called the persistence length. As a consequence of this semi-flexibility, polymers can adopt an infinite number of three-dimensional (3D) arrangements, or conformations. Individual conformations cannot be predicted, much like the positions of individual molecules in a gas cannot be predicted. However, polymer theory can predict statistical quantities, such as the mean distance between the two ends of a polymer chain (Figure 2.3b) or the frequency with which two monomers contact each other (Figure 2.3c). These quantities are predicted to obey scaling laws, which describe how they vary with the number of monomers N or, equivalently, the contour length s of the chain. In the simplest model, the ideal chain, bonds between monomers have random orientations and monomers ignore each other (i.e., neither repel nor attract each other) (Rubinstein and Colby 2003). At equilibrium, in absence of any external constraints or forces, the root mean squared end-to-end distance increases as $\sqrt{s}$ or$\sqrt{N}$ (Figure 2.3b), while the contact frequency decays as $s^{-3/2}$ (Figure 2.3c). The more realistic real chain model accounts for the fact that a polymer cannot self-intersect. This constraint leads to a swelling of the chain and a faster scaling of end-to-end distances (Figure 2.3b). In a confined and crowded volume such as the nucleus, however, this swelling can be counterbalanced by the presence of other chains (and other segments of the same chain). In that case, the scaling laws become similar to the ideal chain up to a distance where monomers behave as if they were no longer part of the same chain, at which point they level off (Figure 2.3b,c).

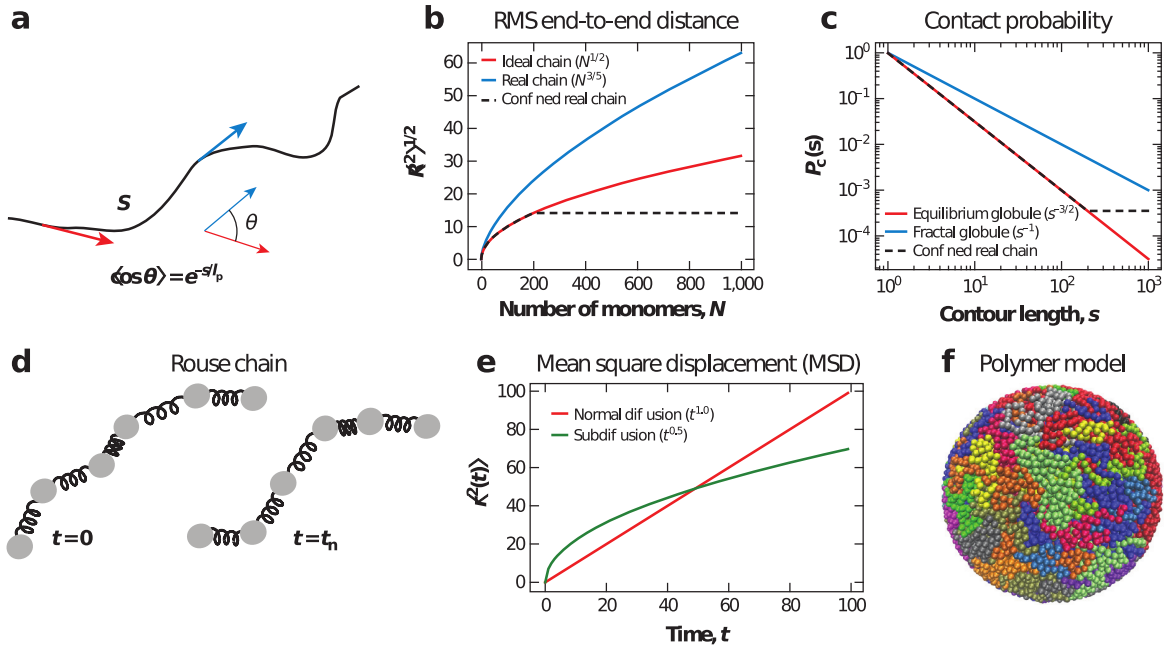


Figure 2.3: **Basic properties of polymers**. **(a)** A polymer is a semi-flexible structure whose rigidity can be defined by the bending persistence length, l_p. The average cosine of Δ, the angle between the tangent vectors at two loci separated by a curvilinear distance, s, decreases as $\exp(-s/l_p)$. **(b)** The root mean squared (RMS) end-to-end distance, $\langle R^2 \rangle^{1/2}$, as a function of the number of monomers, N, for an ideal chain (*red*), a real chain (*blue*), and a chain in confinement (*dashed black*). **(c)** The contact probability, $P_c(s)$, as a function of s between loci for an equilibrated polymer (*red*), a fractal (or crumpled) globule (*blue*) (A. Yu Grosberg, Nechaev, and Shakhnovich 1988), and a confined polymer (*dashed black*). **(d)** A schematic showing the dynamics of a Rouse polymer where monomers are connected by harmonic springs (Rubinstein and Colby 2003). **(e)** Mean squared displacement (MSD), $\langle r^2(t) \rangle$, as a function of time for a freely diffusing monomer (*red*) and a monomer embedded in a polymer chain undergoing Rouse subdiffusion (*green*). **(f)** A snapshot of a molecular dynamics simulation of multiple chromosomes in the nucleus. Panel f courtesy of J.J. Parmar (unpublished manuscript).

Theory can also predict properties of polymer motions. The simplest model, the Rouse model, describes how the random motion of a single monomer is influenced by that of other monomers to which it is connected (Figure 2.3d). According to this model, the mean squared displacement (MSD) of a monomer grows like the square root of time over short timescales (Rosa and Zimmer 2014; Rubinstein and Colby 2003), unlike a free particle, for which the MSD is simply proportional to time (Figure 2.3e).

These predictions are based on a number of important assumptions, most importantly, that the system is in equilibrium and that the polymers consist of identical monomers (homopolymers). Neither of these two assumptions holds true for DNA and chromatin fibers, which are subject to ATP-consuming (i.e., energy-driven) processes such as transcription and replication, and have a nonuniform composition determined by the DNA sequence and epigenetic histone modifications. One might therefore expect the above relationships to utterly fail when applied to chromatin fibers in real biological nuclei. Surprisingly, however, basic homopolymer physics has proven quite effective at explaining some key features of nuclear architecture in a variety of organisms (Rosa and Zimmer 2014). Discrepancies between predictions of basic polymer models and observations are useful because they hint at additional mechanisms of potential biological significance and motivate the development of more realistic and complex models. Examples are heteropolymer simulations, where monomers have distinct types and are subject to different interactions (Figure 2.3f) defined by DNA sequence or epigenetic information, some of which are discussed below (Fudenberg and Leonid A Mirny 2012; Rosa and Zimmer 2014; Bianco et al. 2017; Sazer and Schiessel 2018).

2.3 Nucleosomes and chromatin fiber structure

2.3.1 Beads on a String

The DNA double helix contains approximately three base pairs (bp) per nanometer of length and has a persistence length of $\sim$ 50 nm, i.e., 150 bp, as estimated from in vitro experiments (Wiggins et al. 2006). According to the ideal chain model (Figure 2.3b), an average human chromosome would then have an average size exceeding 50 µm, much larger than typical nuclei. The first level of genome packaging is achieved by nucleosomes, histone octamers whose shape can be approximated by a short cylinder 10 nm in diameter and 5 nm in height. Despite the stiffness of DNA at this scale, 147 bp of DNA wrap around each nucleosome, taking 1.7 turns. This interaction occurs because the positively charged histones are attracted to the negatively charged DNA, resulting in a net free energy gain of $\sim$ 40 k_BT (Mack et al. 2012; Jie Yan et al. 2007). Nucleosomes are spaced by 20–40 bp of linker DNA, such that a stretched array of nucleosomes appears as beads on a string in electron micrographs (Figure 2.4a). Nucleosomes are dynamic and can be repositioned by ATP-dependent chromatin remodelers (Narlikar, Sundaramoorthy, and Owen-Hughes 2013; Z. Zhang et al. 2011). Several equilibrium and non-equilibrium models, with (un)binding and sliding kinetics, have been successful at explaining the positioning and dynamics of nucleosomal arrays (Figure 2.4a) (Kornberg and Stryer 1988; Padinhateeri and Marko 2011; Parmar, Marko, and Padinhateeri 2014).

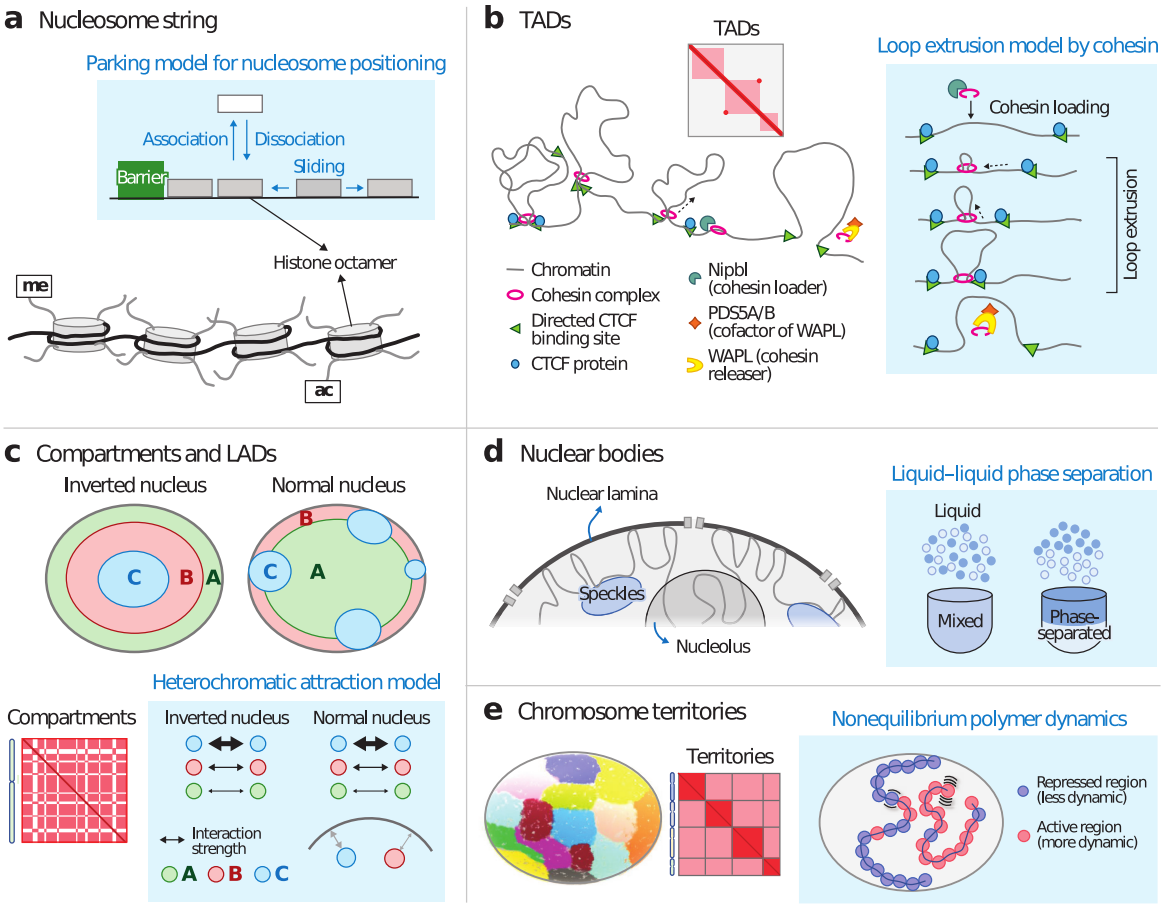


Figure 2.4: **Different levels and mechanisms of chromatin organization**. This figure highlights distinct features of 3D genome architecture and some proposed explanations for their formation (*blue boxes*). **(a)** Chromatin fiber: Wrapping DNA around nucleosomes results in the ∼ 10-nm beads-on-a-string fiber. The positioning of nucleosomes is determined by energetic barriers (*green*), steric hindrance by other nucleosomes, and by ATP-dependent chromatin remodelers that can assemble, disassemble, or slide nucleosomes (*blue box*) (Chereji and Clark 2018; Padinhateeri and Marko 2011; Parmar, Marko, and Padinhateeri 2014). The compaction of nucleosomes in the chromatin fiber is affected by post-translational modifications of histone tails, such as methylation (`me`) or acetylation (`ac`). **(b)** TADs (topologically associated domains) and loops: TADs appear as blocks of higher contact frequency on the Hi-C map diagonal. Loops show up as peaks, often located at TAD corners. The blue box shows chromatin regions undergoing cohesin-mediated loop extrusion (Fudenberg, Abdennur, et al. 2018). In this scenario, cohesin is loaded on DNA by NIPBL and, once loaded, extrudes DNA until reaching a properly oriented CTCF boundary. Cohesin can be released from the chromatin by WAPL and PDS5A/B factors. **(c)** Compartments and LADs (lamina-associated domains): Euchromatin (A, *green*), facultative heterochromatin (B, *red*) and constitutive heterochromatin (C, *blue*) compartments segregate radially in the nucleus, as seen by microscopy and the checkerboard pattern of Hi-C maps. LADs strongly correlate with B compartments. In inverted nuclei (Solovei et al. 2009), the radial organization is reversed, but the contact pattern remains similar. Both can be explained by a heteropolymer model that involves attractions between heterochromatic regions, and additional lamina-dependent interactions with the nuclear envelope to model conventional nuclei (blue box) (Falk et al. 2018) **(d)** Nuclear bodies, including speckles and nucleoli, form membrane-less compartments in the nucleus that appear to be driven by liquid–liquid phase separation (Zhu and Brangwynne 2015). **(e)** Chromosome territories: Distinct chromosomes (shown in color) take separate, non-overlapping positions in the nucleus, with the more transcriptionally active chromosomes preferentially occupying the nuclear center (T. Cremer and M. Cremer 2010). Hi-C maps show high-intrachromosomal and low-interchromosomal contact frequencies. One proposed mechanism of chromosome territory formation (blue box) assumes an activity-dependent dynamics. Red beads represent gene-rich regions (more dynamic), and purple beads represent gene-poor regions (less dynamic) (Ganai, Surajit Sengupta, and Menon 2014).

2.3.2 The Elusive Chromatin Fiber Structure

How are nucleosomes arranged with respect to each other in three dimensions? Early *in vitro* transmission electron microscopy (EM) and X-ray experiments suggested that the beads-on-a-string fiber would fold in a higher-order chromatin fiber of $\sim$ 30 nm in diameter in which nucleosomes are tightly packed together (Bak, Zeuthen, and F. H. Crick 1977; Dorigo et al. 2004; Finch and Klug 1976; Gerchman and Ramakrishnan 1987; Kruithof et al. 2009; Robinson et al. 2006; Schalch et al. 2005; Song et al. 2014; Widom 1985). This structure was thought to arise from the supercoiling of regularly spaced nucleosomes and stabilized by linker histone and electrostatic interactions between the histone tails (Robinson et al. 2006), resulting in DNA compactions far exceeding 100 bp/nm. However, this characteristic structure has not been confirmed in vivo (Fussner et al. 2011; J. C. Hansen et al. 2018; Razin and Gavrilov 2014), and more generally, the intermediate folding of chromatin remains elusive. Recent 3D high-resolution EM data of specifically labeled chromatin (Ou et al. 2017) revealed a complex and irregular folding of the DNA, without 30-nm fiber but with various diameters, ranging from 5 nm to 24 nm (Figure 2.5f). Recent fluorescence super-resolution imaging data are also consistent with the absence of 30-nm fibers and a polymorphic structure (Nozaki et al. 2017; Ricci et al. 2015).

Modeling studies (Figure 2.5g) can also provide information about the chromatin fiber structure and estimates of its compaction and rigidity based on contact frequency measurements from Hi-C techniques or its ancestors (Figure 2.5a) (Dekker et al. 2002) and distances between loci from imaging. In yeast, an analysis using a whole nucleus simulation inferred an average compaction of 53–65 bp/nm– also arguing against a 30-nm fiber– and a persistence length of 52–85 nm, i.e., $\sim$ 3–6 kb (Arbona et al. 2017). Based on these numbers, the ideal chain model would predict a human chromosome size of $\sim$ 15 µm. This is much larger than the actual size of chromosomes, calling for other explanations, as discussed in section 2.6.

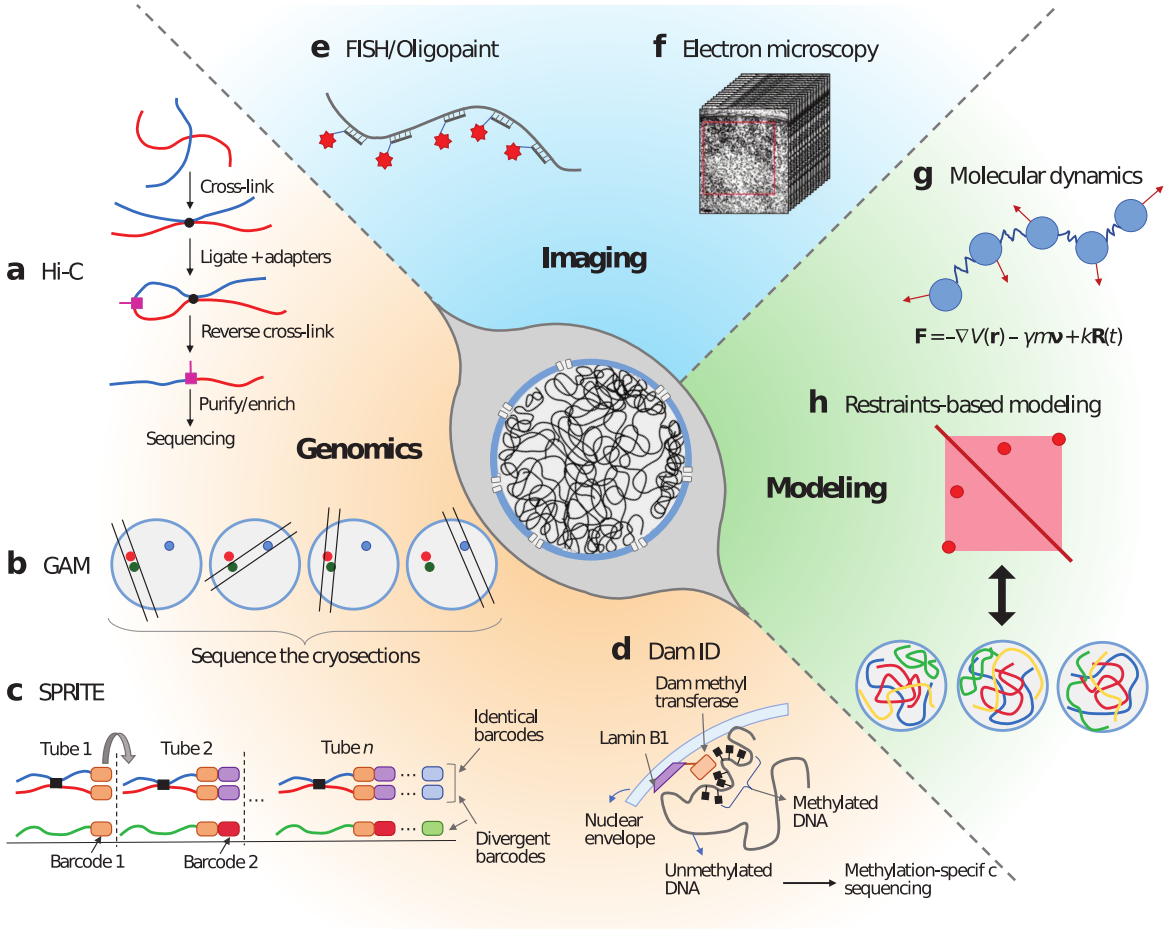


Figure 2.5: **Techniques to study chromatin organization**. They fall into three main categories: genomics *(a–d)*, imaging *(e,f)*, and modeling *(g,h)*. **(a)** In Hi-C (Lieberman-Aiden et al. 2009), different chromatin regions that are in close spatial proximity are cross-linked, fragmented, ligated, and marked with adapters (*pink*). Fragments are then reverse cross-linked, purified, sequenced, and mapped to their genomic locations, yielding genome-wide contact frequency matrices. **(b)** In GAM (genome architecture mapping), nuclei are cryo-sectioned into thin slices and their DNA is sequenced (Beagrie et al. 2017). Analysis of locus co-occurrence in many sections allows one to infer proximity, including multiway interactions, without ligation. **(c)** In SPRITE (split-pool recognition of interactions by tag extension), DNA and RNA fragments are barcoded in a sequential manner that allows one to detect both DNA–DNA and DNA–RNA associations by sequencing (Quinodoz et al. 2018). **(d)** In DamID (DNA adenine methyltransferase identification), chromatin regions close to the nuclear lamina are marked by the Dam methyl transferase and are mapped genome-wide by sequencing (Guelen et al. 2008; van Steensel and Henikoff 2000). **(e)** DNA FISH (fluorescence in situ hybridization) methods allow one to visualize targeted chromatin domains or entire chromosomes in single cells (Beliveau et al. 2015; Bolzer et al. 2005; S. Wang et al. 2016). **(f)** Electron microscopy, in combination with DNA-specific labeling, can reveal nanometer-scale 3D chromatin structures in frozen samples (Ou et al. 2017). **(g)** Molecular dynamics simulations can model time-dependent changes in chromosome configurations by computing the displacement of monomers based on internal and external forces (Arbona et al. 2017; Rosa and Everaers 2008; Rosa and Zimmer 2014). Such models can predict contact frequencies and average locus positions from a relatively small number of assumptions. **(h)** Inverse modeling approaches typically use Hi-C data to reconstruct a population of 3D structures consistent with the data (Kalhor et al. 2011; Rosa and Zimmer 2014).

2.4 Topologically Associated Domains and loops

2.4.1 Topologically Associated Domains, Loops, and Stripes

While EM and light microscopy have allowed insights into the structure of chromatin and chromosomes at small (nucleosome-level) and large (nuclear) scales (see section 2.6), intermediate scales of chromatin folding have long remained obscure. A major milestone enabled by Hi-C was the discovery that intrachromosomal contact frequency matrices display squared blocks of higher frequencies along the diagonal, reflecting regions within which contacts occur more frequently than with any other parts of the genome, and where the average contact frequencies decay slower than the genome-wide average (Dixon et al. 2012; Nora, Lajoie, et al. 2012) (Figures (Figures 2.4b), 2.6a, and 2.7). These domains were called topologically associated domains (TADs) (note that in this context "topological" does not carry its physical meaning). Although dependent on the algorithm used to define them, the number of TADs in human cells has been reported as $\sim$ 10,000, with a median size of $\sim$ 200 kb (S. S. Rao et al. 2014). While TADs were first characterized in population-averaged Hi-C maps, later imaging experiments support their presence as physical units in single nuclei (Szabo et al. 2018). Another prominent feature uncovered by Hi-C are chromatin loops, identified as peaks in the contact maps (Figure 2.6b). A large fraction of TADs have such peaks at their corners and conversely, many loops are associated with TADs (S. S. Rao et al. 2014). With sufficient sequencing coverage, smaller TADs can be found nested within bigger ones, and these sub-TADs tend to share common contact peaks. In addition, many TADs feature stripes at their edges, indicating that contacts between a locus at the boundary and all other loci within the TAD are more frequent than between random pairs of loci within the TAD (Figure 2.6c) (Vian et al. 2018).

2.4.2 Loop Extrusion

What mechanism can explain the formation of loops, TADs, and stripes? Arguably the simplest scenario is that contacts between distant sites first occur because of random collisions between monomers in a polymer undergoing thermal fluctuations. Most of these random contacts will be short-lived; however, if two colliding loci are bound by molecular factors that can form longer-lived interactions, these will stabilize contacts and create local peaks in the Hi-C matrix. Indeed, the boundaries of TADs are strongly enriched in specific proteins, most notably the insulator protein CTCF and cohesin (S. S. Rao et al. 2014). However, the random collision model should also generate contact enrichments with other domains on the same chromosome, and even with other chromosomes, leading to off-diagonal blocks in the contact maps –a prediction that is not borne out by Hi-C data. This model also fails to explain another key observation, namely that the boundaries of TADs correlate strongly with converging CTCF sites, i.e., motifs oriented towards each other, while the three other possible orientations of these motifs are strongly disfavored (S. S. Rao et al. 2014). Finally, how TADs with only one stripe can arise in this scenario is similarly unclear.

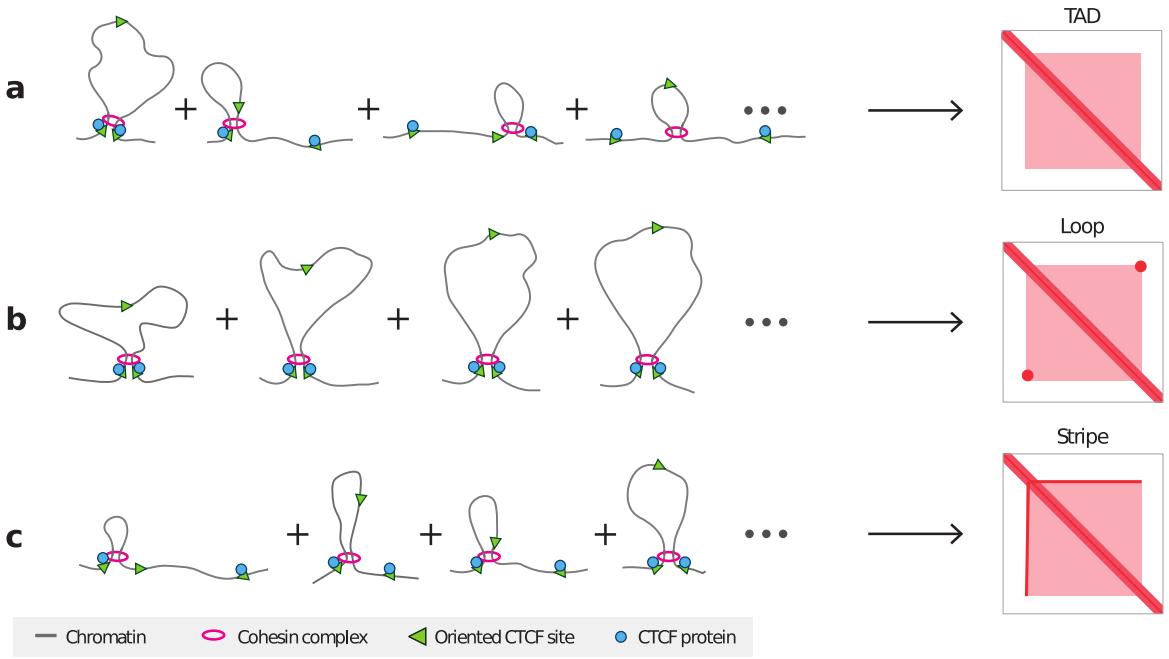


Figure 2.6: Loop extrusion scenarios and the resulting contact maps. (*left*) Multiple conformations of the same chromatin region in different cells, and (*right*) the resulting population-averaged contact frequency maps (i.e., the expected Hi-C maps). **(a)** TAD (topologically associated domain) formation: Loops extruded bidirectionally by cohesin landing at random positions generate an enrichment of contacts within a domain (defined by converging CTCF sites), as reflected by a square on the main diagonal of the contact map. **(b)** Loop extrusion stops at converging CTCF sites, giving rise to a contact frequency peak at the TAD corner. **(c)** Cohesin landing near a CTCF site and extruding chromatin unidirectionally yield a stripe at the TAD boundary.

By contrast, all of these observations can be simultaneously recapitulated by a very different mechanism known as loop extrusion (Alipour and Marko 2012; Fudenberg, Imakaev, et al. 2016; Nasmyth 2001; Nuebler, Fudenberg, Imakaev, Abdennur, and Leonid A. Mirny 2018; S. S. Rao et al. 2014; Sanborn et al. 2015) (Figures 2.4b and 2.6). According to this model, cohesin rings land on chromatin and actively pull out (extrude) the DNA until they fall off or encounter obstacles such as CTCF bound sites (with the right orientation) or other cohesin complexes. Although assuming some ad hoc parameters (e.g., an average processivity of $\sim$ 200 kb, a 10% permeability of CTCF boundaries, and uni- or bidirectional movements), molecular dynamics simulations of loop extrusion (Fudenberg, Imakaev, et al. 2016; Nuebler, Fudenberg, Imakaev, Abdennur, and L. Mirny 2017; Sanborn et al. 2015; Vian et al. 2018) are remarkably successful at explaining almost all the experimental evidence mentioned above, as well as the effect of several experimental perturbations summarized below (Fudenberg, Abdennur, et al. 2018).

2.4.3 Molecular and Energetic Determinants of Topologically Associated Domains

Several lines of evidence directly or indirectly support the loop extrusion model and indicate the role of key molecular determinants of TAD formation. Induced degradation of the cohesin subunit Rad21 leads to the complete disappearance of TADs in less than 40 min, but restoration of cohesin recovers TADs within 15–40 min, demonstrating the crucial role of cohesin in TAD formation (Figure 2.7a) (S. S. P. Rao et al. 2017; Wutz et al. 2017). Deletion of NIPBL, a protein that loads cohesin on the DNA, results in similar effects (Figure 2.7c) (Schwarzer et al. 2016), implying that cohesin must be loaded repeatedly on chromatin to maintain TADs. While $\sim$ 90% of cohesin binding sites coincide with CTCF sites, only 29% coincide with Nipbl sites, and only $\sim$ 11% of CTCF sites coincide with Nipbl sites, indicating that cohesin is loaded outside of CTCF sites and, once loaded, translocates very

fast from the loading site to CTCF sites (Busslinger et al. 2017; Vian et al. 2018; Zuin et al. 2014).

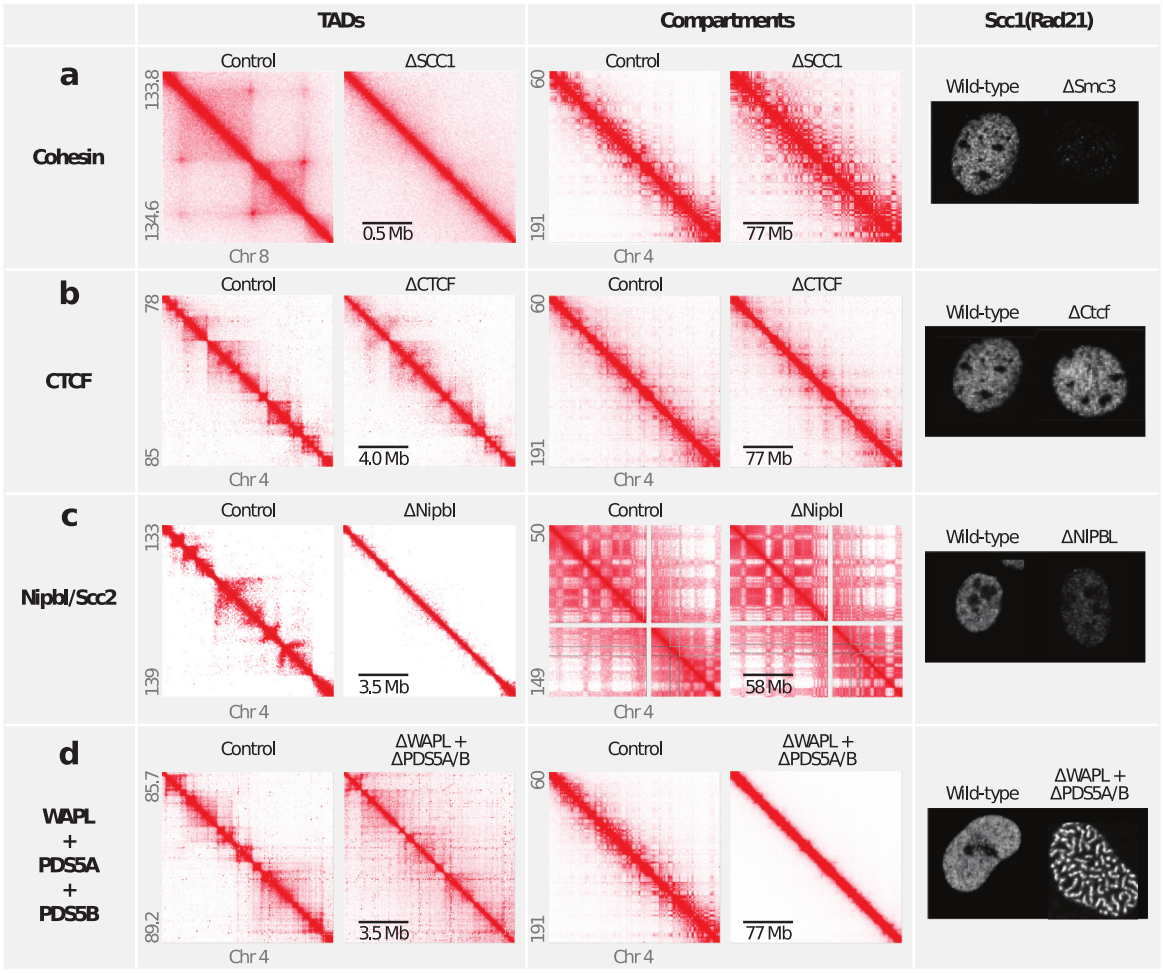


Figure 2.7: Molecular determinants of chromatin domains, showing the effect of depleting or deleting individual architectural proteins on Hi-C contact maps at the scale of TADs (topologically associated domains) and compartments and on microscopy images of the cohesin component RAD21 (SCC1). **(a)** Depletion of SCC1 removes TADs and strengthens compartments. Imaging shows no Scc1 signal when the cohesin subunit Smc3 is deleted, indicating depletion of the cohesin complex. **(b)** CTCF depletion blurs TAD boundaries but has no effect on compartments. Imaging shows no visible change of Rad21. **(c)** Degradation of the cohesin loader NIPBL has similar effects on TADs and compartments as degradation of cohesin and leads to a strongly decreased Rad21 signal. **(d)** Degradation of WAPL along with its cofactors PDS5A and PDS5B results in the appearance of new loop peaks, enlargement of TADs, and removal of compartment structures. Imaging of Rad21 shows condensed chromatin structures similar to those seen during mitosis. Depletion of WAPL or the two cofactors alone results in similar, but less pronounced, effects (not shown). Hi-C data from (a) HCT116 (human) (S. S. P. Rao et al. 2017), (b,d) HeLa (human) (Wutz et al. 2017), and (c) hepatocytes (mouse) (Schwarzer et al. 2016). Visualization of contact maps was done with Juicebox (Durand et al. 2016). Microscopy images are of (a,b) embryonic fibroblasts (mouse) (Busslinger et al. 2017), (c) HAP1 cells (human) (Haarhuis et al. 2017), and (d) HeLa (human) (Wutz et al. 2017).

Conversely, deletion of WAPL or PDS5A and PDS5B, proteins that cooperate to release cohesin from the DNA, yields an enlargement of TADs by more than 200 kb and a proliferation of loops (Figure 2.7d) (Haarhuis et al. 2017; Wutz et al. 2017). In absence of WAPL, cohesin can travel far distances and even bypass CTCF, resulting in a loss of interphase chromatin organization characterized by condensed mitotic-like chromatin termed vermicelli (Figure 2.7d) (Busslinger et al. 2017; Haarhuis

et al. 2017; Tedeschi et al. 2013; Wutz et al. 2017). These observations suggest that turnover of cohesin is necessary for proper chromatin organization during interphase and that, if left on chromatin, cohesin keeps extruding longer lengths of DNA, resulting in very condensed chromosomes.

The degradation of CTCF proteins does not remove TADs but makes their boundaries fuzzier, in accordance with the above idea that CTCF is not directly involved in TAD formation but rather in defining the boundaries (Figure 2.7b). Disruption or flipping of CTCF binding sites by genome editing results in changes of TAD boundaries (Nora, Lajoie, et al. 2012), e.g., fusion of consecutive TADs, in excellent agreement with simulation predictions (Sanborn et al. 2015). Moreover, disruption of Ctcf (in mouse) results in several new cohesin peaks at the active transcription sites (Busslinger et al. 2017), suggesting that absence of CTCF allows cohesin to travel longer distances until it finds another roadblock, which in this case could be active transcription. How CTCF or active transcription sites block cohesin is still an open question.

While some models assume that extrusion relies on an energy-driven activity of cohesin as a molecular motor (Fudenberg, Imakaev, et al. 2016; Sanborn et al. 2015), others propose extrusion without such activity, powered by either transcriptionally induced super-coiling or even mere thermal diffusion (Pereira et al. 2018; Racko et al. 2018). Experimentally depleting the cells of ATP shifts the genome-wide distribution of cohesin away from CTCF sites and toward Nipbl binding sites, and prevents the reformation of TADs when restoring cohesin levels after induced degradation. These experiments support an energetic requirement for cohesin translocation from the loading sites and for TAD formation (Vian et al. 2018).

Further independent evidence in favor of the extrusion model comes from *in vitro* single-molecule imaging experiments showing that the human cohesin complex can translocate on DNA and bypass single nucleosomes and DNA-bound proteins, but not CTCF (Davidson et al. 2016). However, in this study translocation was independent of ATP, whereas in yeast, cohesin loading is ATP-dependent (Murayama, Samora, et al. 2018; Murayama and Uhlmann 2014; Murayama and Uhlmann 2015). Another recent single-molecule experiment with the related yeast condensin complex showed fast, unidirectional, ATP-dependent loop extrusion ($\sim$ 1500 bp/s, step size $\sim$ 50 nm) on linear DNA (Ganji et al. 2018). The dynamic nature of loops is also supported by single-molecule tracking of cohesin and CTCF *in vivo* (Anders S. Hansen, Pustova, et al. 2017). Further experimental studies are required to definitively establish active loop extrusion as the mechanism of TAD formation.

Finally, although the role of cohesin in TAD formation is supported by many Hi-C experiments, a recent imaging study showed that even in the absence of cohesin, TAD-like structures remain present in single cells; however, their boundaries become randomized along the genome and no longer preferentially associate with CTCF sites (Bintu et al. 2018). Moreover, the TAD-like domains were reestablished after mitosis in the absence of cohesin. These new results question the role of cohesin in defining chromatin domains in single cells and call for more investigations.

2.4.4 Functional Role of Topologically Associated Domains

A key property of TADs is their high degree of conservation between cell types and species (Dixon et al. 2012; S. S. Rao et al. 2014). In line with this, TADs and loop extrusion are thought to be both associated and crucial to several biological processes.

1. **Transcription.** In terms of gene expression, TADs can be seen as highly functional units: Genes within the same TAD tend to be co-regulated (Zhan et al. 2017) and loops correlate with enhancer–promoter interactions (S. S. Rao et al. 2014). Alterations of TADs can lead to abnormal expression patterns. For instance, disruption of a single TAD boundary in mice was sufficient to induce polydactyly, a severe developmental malformation (Lupiáñez et al. 2015). This effect was explained by the induction of contacts between an enhancer and the promoter of a developmental gene that were previously insulated from each other by the TAD boundary. Similarly, it was shown that loss of CTCF at a TAD boundary in patient-derived cells leads to aberrant contacts of a constitutive enhancer with an oncogene and hyper-activation of its expression, resulting in

increased cell proliferation (Flavahan et al. 2016). Such evidence underlies the view that TADs provide a means to constrain the action of enhancers to a small number of promoters. However, this immediately raises the question as to how the relatively moderate contact insulation afforded by TADs –contacts within TADs are on average only $\sim$ 2–3 times more frequent than across TADs (Merkenschlager and Nora 2016)– can explain the fact that gene expression essentially follows an all-or-nothing behavior depending on whether both an active enhancer and a promoter reside in the same TAD (Flavahan et al. 2016; Lupiáñez et al. 2015). One possible solution to this conundrum might be the unidirectional extrusion mentioned above to explain contact stripes (Figure 2.6c) (Vian et al. 2018). If one of the loop anchors is fixed at a promoter, DNA extrusion can bring this locus in contact sequentially with the entire TAD domain, including all potential enhancer sequences, without having to rely on random 3D collisions. This process might greatly increase the frequency of interactions between pairs of enhancers and promoters that share a TAD (and a stripe) compared to pairs that do not.

Another potential answer might come from the timescales of promoter activation by enhancer interactions. In the traditional view of mammalian gene expression, a physical contact between the enhancer sequence bound by activating transcription factors and the transcription pre-initiation complex is required to initiate transcription, and many lines of evidence support this model (Deng et al. 2014). However, it is unclear whether this physical contact is actually required for the RNA polymerase to initiate transcription, or whether it simply potentiates the polymerase, enabling it to initiate transcription later, even in absence of enhancer–promoter contact. In *Drosophila melanogaster*, enhancer–promoter contacts and transcription seem to be highly synchronized, and more generally, FISH (fluorescence in situ hybridization) experiments (Figure 2.5e) clearly demonstrate a lower enhancer–promoter distance in active genes than in inactive genes, arguing for a direct link between contacts and transcription (Hongtao Chen et al. 2018; B. Lim et al. 2018). On the other hand, both cohesin and CTCF depletion only show minor effects on gene expression over a six-hour window, suggesting that on a population scale, transcription is already potentiated and proceeds as before, even in the absence of TAD delimitation (Nora, Goloborodko, et al. 2017; S. S. P. Rao et al. 2017). These findings are corroborated by recent single-locus imaging in mouse embryonic stem cells, in which no correlation was found between enhancer–promoter distance and transcript production (Alexander et al. 2018). It also remains unclear to what extent specific histone modifications could mediate this potentiation.

While there is evidence of a role for TADs in regulating transcription, there is also evidence of a reverse role of transcription in TAD organization, such as the fact that, in *D. melanogaster*, a large portion of TAD boundaries are enriched in active RNA polymerase instead of CTCF, and that in those cases, TADs are much better defined by active histone marks than by CTCF (Busslinger et al. 2017; Ulianov et al. 2016). In general, however, the interplay between TADs, loops, and transcriptional activation is far from understood and remains to be further elucidated.

2. **Replication.** To enable the replication of the entire human genome in a few tens of minutes, DNA polymerases initiate replication in parallel at several points along chromosomes, leading to replication domains that grow and merge until the entire chromosome is replicated. Surprisingly, TADs coincide almost perfectly with replication domains (Pope et al. 2014), suggesting that TADs might also orchestrate replication, in addition to transcription. This co-localization of TADs and replication origins has recently been used to visualize TAD dynamics in live cells (W. Xiang et al. 2018). One recent study found that the progressive establishment of TADs in early zygote development was prevented by replication inhibition as opposed to transcription inhibition (Ke et al. 2017), but another study found that loop domains can reform after restoration of cohesin despite inhibition of replication (Vian et al. 2018). Clearly, more work is needed to address the mechanistic links between replication and TAD formation.

2.5 Chromatin compartments

2.5.1 A and B Compartments and Lamina-Associated Domains

EM images of interphase nuclei typically show dense heterochromatic regions near the nuclear envelope and around nucleoli, with less compact euchromatin in the center of the nucleus and beneath nuclear pores. Notable exceptions to this general rule are retinal cells of nocturnal mammals, which display an inverted nuclear architecture where euchromatin relocates to the periphery while constitutive heterochromatin resides at the nuclear center (Solovei et al. 2009). A new view of this partitioning of the genome resulted from the identification of chromatin compartments in the first Hi-C study (Lieberman-Aiden et al. 2009). Compartments appear as a checkerboard pattern in the Hi-C map after correcting for the average dependence of contact frequencies on genomic distance (the pattern becomes more apparent when computing the corresponding correlation matrix) (Figures 2.4c and 2.7) (Lieberman-Aiden et al. 2009). This pattern and an eigenvector analysis suggest that the entire genome is partitioned to first order in alternating regions of a few megabases belonging to two compartments (called A and B) –subsequent Hi-C experiments with higher resolution further refined this partitioning into six sub-compartments (S. S. Rao et al. 2014). Unlike for TADs, a locus belonging to the A compartment exhibits enriched contacts with other loci from the A compartment throughout the genome but has less frequent contacts with loci from the B compartment (and vice-versa). Comparisons with ChIP-Seq (chromatin immunoprecipitation and sequencing) data show a strong correlation of A regions with transcriptionally active histone marks and decondensed chromatin, while B regions correlate with inactive histone marks and dense regions, thereby providing a new definition of euchromatin and heterochromatin. Imaging experiments have later confirmed the existence of A and B compartments (as well as TADs) with different levels of compaction in single cells of *D. melanogaster* (Lieberman-Aiden et al. 2009; S. Wang et al. 2016). The B compartment identified by Hi-C also exhibits very high correlation with lamina-associated domains (LADs), chromatin regions in contact with the nuclear envelope, as identified by DamID (DNA adenine methyltransferase identification) (Figure 2.5d) (Guelen et al. 2008; van Steensel and Belmont 2017).

2.5.2 Compartmentalization Mechanisms

Although A and B compartments were identified several years before TADs, our understanding of the mechanisms underlying their segregation is comparatively less advanced. Most explanatory models derive from the observation that A and B compartments are enriched for specific histone modifications (Lieberman-Aiden et al. 2009). These modified histones can act as scaffolds for other proteins capable of interacting with more than one histone or proteins exhibiting a high level of self-interaction, both of which can potentially mediate phase separation (Erdel and Rippe 2018). Alternatively, compartment segregation might result from associations to (at least) two types of anchors: the nuclear lamina, which contains several proteins known to interact with modified histones, and nuclear speckles, which are located more centrally (see section 2.6.2) (W. Chen et al. 2018; Yu Chen et al. 2018; van Steensel and Belmont 2017). Other proposed models invoke differences in chromatin dynamics or transcriptional activity (Ganai, Surajit Sengupta, and Menon 2014).

In recent studies (Falk et al. 2018; Nuebler, Fudenberg, Imakaev, Abdennur, and L. Mirny 2017; Pereira et al. 2018), chromosomes were modeled as heteropolymers (Figure 2.4c) partitioned into three compartments based on the Hi-C maps (A, B, and constitutive heterochromatin, termed C). The simulations recovered the inverted nuclear architecture mentioned above as well as the compartment organization by assigning realistic interactions between the three compartments, e.g., an attractive energy potential of 0.5 k_BT for interactions between B compartment monomers of 30 kb. Keeping the same model but adding interactions of B and C monomers with the nuclear envelope recovers the conventional nuclear architecture. This suggests that the compartment segregation may be explained by attractive interactions between heterochromatin regions, rather than between euchromatin regions, and is unrelated to tethering at the nuclear lamina, and that the default organization of chromatin (in

absence of tethering with the lamina) is the inverted one (Falk et al. 2018). Although the model does not specify the molecular nature of the assumed heterochromatic interaction, one plausible candidate is heterochromatin protein 1 alpha (HP1α), which has been shown to undergo phase separation both *in vitro* (A. G. Larson et al. 2017) and *in vivo* (Strom et al. 2017). Clearly, more work is needed to expand on such early results and fully address the molecular and physical mechanisms that underlie chromatin compartmentalization at the megabase scale.

2.5.3 Functional Implications

The partitioning of the genome in compartments and LADs correlates with functional processes: B compartments and LADs tend to be transcriptionally inactive and late replicating, and A compartments tend to be transcriptionally active and early replicating. A central question is to what extent these correlations reflect causes or consequences. Segregation provides spatially separated sub-compartments that are amenable to different types of reactions with different kinetics (Woringer and Darzacq 2018). For example, the high density of heterochromatin, as evidenced by, e.g., recent EM data (Ou et al. 2017), might prevent the assembly of the multiple-megadalton pre-initiation complex and thereby silence gene expression. Consistent with this view, experimental tethering of genes to the nuclear membrane can lead to their transcriptional repression (Akhtar et al. 2013; Reddy et al. 2008). However, experimentally induced chromatin decondensation by recruitment of an acidic peptide did not lead to transcriptional activation, arguing against a direct link between chromatin compaction and gene expression (Pierre Therizols et al. 2014). Thus, a complete picture is missing and future research should further explore the causal relations between compartments, LADs, and gene expression.

2.6 Nuclear-scale organization

2.6.1 Chromosome Territories

After mitosis, chromosomes decondense, and owing to their large size, limited compaction, and flexibility, they would be expected to intermingle and fill out the entire nucleus (Rosa and Zimmer 2014). However, microscopy has shown that the interphase nuclei of many mammalian cells are partitioned into largely disjoint chromosome territories (Figure 2.4e) (T. Cremer and M. Cremer 2010) whose radial positions display statistical preferences. For instance, gene-rich chromosomes tend to occupy central positions while gene-poor chromosomes tend to be more peripheral (Bolzer et al. 2005; Croft et al. 1999). In yeast, by contrast, chromosomes strongly intermingle despite their much smaller size (Berger et al. 2008; P. Therizols et al. 2010).

1. **Mechanisms.** What might explain the different organization of chromosomes in these organisms? One line of explanation is based on the timescale needed for chromosomes to relax after mitosis. Because topological constraints (the assumed inability of distinct polymer chains to cross each other) increase the relaxation time of polymers as the third power of their length, this time was predicted to be much longer than the cell cycle for mammalian genomes, but not for yeast. Thus, human chromosomes might simply not have enough time to equilibrate and mix, and interphase territories could reflect the individuality of mitotic chromosomes (Rosa and Everaers 2008). Support for the prediction that mammalian chromosomes are in an out-of-equilibrium state came from the first Hi-C study, which indicated that the genome-wide average contact frequency is inversely proportional to genomic distance, s (that is, $P_c(s) \propto s^{-1}$) (at least between $\sim$ 500 kb and $\sim$ 5 Mb), in contrast to the $P_c(s) \propto s^{-3/2}$ scaling expected for equilibrium (Figure 2.3c) (Lieberman-Aiden et al. 2009). A fractal (or crumpled) globule model, where the polymer remains untangled in contrast to an equilibrated model, was proposed to explain this scaling (A. Yu Grosberg, Nechaev, and Shakhnovich 1988). Conversely, the prediction that yeast chromosomes are at equilibrium is supported by the fact that a Brownian dynamics simulation can successfully account for imaging and Hi-C data, including the contact frequency scaling (Arbona et al. 2017; H. Wong et al. 2012).

However, despite the attractiveness of these generic and largely parameter-free models, several questions remain. First, the action of topoisomerase II—which cuts both strands of the DNA double helix—was ignored in these simulations but is expected to strongly reduce equilibration time by relaxing topological constraints. Whether chromosome territory formation can be explained with realistic modeling of topoisomerase II action remains to be seen, particularly for post-mitotic cells such as neurons. Second, the incomplete relaxation model (Rosa and Everaers 2008) assumed homopolymers and ignored the presence of both compartments, TADs, and differences in transcriptional activity, which other models predict to affect chromosome positioning (Ganai, Surajit Sengupta, and Menon 2014). It will therefore be interesting to revisit the formation of chromosome territories in the context of interactions –e.g., electrostatic (Strom et al. 2017)– between compartments, loop extrusion, and differential dynamics, all of which affect the size of interphase chromosomes, even at steady state.

2. **Functional implications.** Much as for A/B chromatin compartments, the functional relevance of chromosome territories remains unclear. It has been proposed that chromosome territories facilitate chromosome condensation prior to mitosis (Rosa and Everaers 2008). Moreover, because the spatial proximity of loci or chromosomes correlates with increased translocations, a hallmark of cancer cells, it has been proposed that the organization in territories acts to minimize interchromosomal rearrangements (Branco and Pombo 2006; Nikiforova 2000).

2.6.2 Nuclear Bodies

Nuclear bodies are sub-compartments of the nucleus that lack a bona fide membrane. They are usually visible by phase contrast microscopy as spheroid, often dynamic structures comprising a dense aggregate of proteins, RNA, and potentially many other macromolecules and are associated with specific functions such as transcription of certain genes, splicing, DNA damage repair, etc. (Figure 2.4d). The prototypical nuclear body is the nucleolus, the site of ribosomal RNA biogenesis (Boisvert et al. 2007). Since its discovery, many other ubiquitous or species-specific nuclear bodies have been identified, including nuclear speckles, Cajal bodies, promyelocytic leukemia bodies, histone locus bodies, and paraspeckles (Matera 1999). The nucleation or morphology of nuclear bodies intimately depends on their activities, for example, inhibition of ribosomal DNA transcription strongly reduces the nucleolar volume (P. Therizols et al. 2010) and can be regulated by post-translational modifications such as phosphorylation of RNA polymerase II (Kwon et al. 2013).

1. **Mechanisms.** The mechanisms underlying nuclear body formation are under active investigation, with increased attention on liquid–liquid phase separation (Figure 2.4d) (Hyman, C. A. Weber, and Jülicher 2014). Several examples of phase separation inside the nucleus have been reported, often mediated either by proteins bearing unstructured domains (A. G. Larson et al. 2017; Strom et al. 2017), RNAs (Feric et al. 2016), or other small molecules (Altmeyer et al. 2015; Patel, Malinovska, et al. 2017). For example, the C-terminal domain of RNA polymerase II has been shown *in vitro* to perform reversible and regulatable phase separation (Kwon et al. 2013), and recent evidence suggests that similar mechanisms exist *in vivo* as well (Boehning, Dugast-Darzacq, Rankovic, Anders S. Hansen, T. Yu, et al. 2018). Another example mentioned above is HP1α, which was observed to form droplets in live-cell fluorescence microscopy of *D. melanogaster* (Strom et al. 2017). In addition, de-phosphorylation promotes droplet disassembly, suggesting a mechanism by which demixing can be regulated. A challenge for coming years is to integrate quantitative models of phase separation with the additional constraints imposed by the physics of polymers (Rosa and Zimmer 2014).

2. **Functional implications.** Although lacking a strict membrane, nuclear bodies can sequester some molecules and exclude others, thereby acting as chemical reactors to catalyze specific reactions, decoupling them from different pathways. An interesting property of nuclear bodies created by phase separation is that their very existence can depend on whether their molecular

constituents exceed a concentration threshold. This nonlinear behavior could endow the cell with a switch-like response to external stimuli (Zhu and Brangwynne 2015).

Another potential function, or consequence, of nuclear body formation is the very organization of chromatin itself. Indeed, it has been shown that chromatin organizes non-randomly around nuclear bodies. While it is difficult to chart sequences in close proximity to a nuclear body with traditional methods, novel techniques have recently been developed that dispense of proximity ligation to probe for DNA sequences located within much larger distances of other DNA or RNA sequences (Figure 2.5c,d) (Quinodoz et al. 2018) or even a specific protein (W. Chen et al. 2018; Yu Chen et al. 2018). These techniques reveal a key role for nuclear speckles, and either LADs (W. Chen et al. 2018; Yu Chen et al. 2018) or nucleoli (Quinodoz et al. 2018), in overall chromatin organization and suggest a plausible mechanism for locus positioning that merits further investigation.

2.7 Interactions between levels

A natural question is how the different levels of organization discussed above interact with each other, either cooperatively or antagonistically. Since chromosomes are single connected structures, one might surmise that these levels are all closely interleaved and hard to disentangle. Therefore, studies that peel off individual layers of organization are particularly instructive. One striking recent example is the experimentally demonstrated partial decoupling of TADs and compartments (S. S. P. Rao et al. 2017; Schwarzer et al. 2016). In absence of cohesin or its loader Nipbl, the segregation of A and B compartments not only persists but is actually strengthened (Figure 2.7a,c): Boundaries between compartments become sharper, and their correlations with histone modifications increase. Conversely, deleting the cohesin release factors (WAPL or PDS5A/B) leads to an enlargement of TADs and to a destruction of A/B compartments (Figure 2.7d) (Wutz et al. 2017). These experiments indicate that the formation of TADs and the formation of compartments rely on distinct mechanisms that partly counteract each other. A similar antagonism is at work between the mechanisms that tend to keep chromosomes in distinct territories and those that create A/B compartments, which tend to mix regions of different chromosomes belonging to the same compartment. Another salient finding is the above mentioned study showing that A/B compartments subsist in the absence of tethering to the nuclear lamina, hence decoupling compartmentalization from LADs (Falk et al. 2018).

Although a general model linking all four levels of chromatin organization, from nucleosomes to the entire nucleus, is still lacking, such experiments and modeling approaches point to future unifying frameworks. Further studies are needed to disentangle the competing forces that shape chromatin architecture in the context of transcription and other functional processes.

2.8 Dynamic organization of the nucleus

The 3D architecture of the genome is by no means static, since chromatin is in constant motion, as expected from basic polymer dynamics (Figure 2.3e) and evidenced by live-cell microscopy in yeast (Herbert et al. 2017; Heun 2001) and mammalian cells (B. Chen et al. 2013; S. Chong, Dugast-Darzacq, Z. Liu, Dong, G. M. Dailey, et al. 2018). By itself, this mobility leads to stochastic variations in chromosome configurations in cell populations. The variability in chromatin organization can also be analyzed in fixed cells using imaging (Boettiger et al. 2016) or single-cell Hi-C methods (Nagano, Lubling, Stevens, et al. 2013; Ramani et al. 2017) or even can be inferred from population Hi-C by computational reconstruction methods (Figure 2.5h) (Kalhor et al. 2011). In addition to the dynamics over short timescales in interphase cells, chromatin organization changes dramatically during each cell division and is altered during differentiation, as briefly discussed below.

2.8.1 Chromatin Organization During the Cell Cycle

It has long been known from light microscopy that chromosomes undergo major structural changes during mitosis, when the sister chromatids condense and align on a metaphase plate before being pulled apart into the two future daughter cells. Hi-C in mammalian and yeast genomes has provided new insights into how chromosome structure changes during the cell cycle (Gibcus et al. 2018; Kakui et al. 2017; Lazar-Stefanita et al. 2017; Nagano, Lubling, Várnai, et al. 2017; Naumova et al. 2013; Schalbetter et al. 2017). A first analysis of synchronized cells showed that both TADs and compartments remain mostly unchanged throughout interphase, with only moderate changes in strength and few changes in boundaries, but are completely lost during mitosis. Mitotic chromosomes assume a universal folding structure independent of cell type that was well described by an array of random $\sim$ 100-kb-long loops (Naumova et al. 2013). In a subsequent study (Gibcus et al. 2018), the time line was further refined to analyze the successive stages of mitosis every few minutes and the role of condensin proteins in shaping mitotic chromosomes. Especially noteworthy was the appearance of a second diagonal in the Hi-C matrix at prometaphase, indicative of a helical chromosome structure, which disappeared upon degradation of condensin II. To explain their Hi-C and imaging data, the authors developed sophisticated polymer models where condensin II creates progressively bigger loops of up to $\sim$ 700 kb by the same extrusion process discussed for cohesin in Section 4.2, in agreement with single-molecule experiments showing condensin-mediated extrusion (Ganji et al. 2018). These big loops are further folded into smaller ($\sim$ 80 kb) loops extruded by condensin I. The condensin II loop anchors are assumed to form a scaffold that adopts a helical structure from which the nested loops emanate radially in a spiral staircase arrangement. Interestingly, a super-resolution imaging study of condensin I and II in mitotic chromosomes provides independent support for the nested loop arrangement (although the helicity could not be ascertained), with similar quantitative estimates of loop sizes (Walther et al. 2018). Importantly, even though deletion of condensin I and II resulted in some morphological changes in mitotic chromosomes, chromosomes remained condensed, suggesting the involvement of some other unknown proteins in mitotic condensation (Gibcus et al. 2018).

2.8.2 Chromatin Reorganization During Early Development

Although interphase chromatin organization is completely lost during mitosis, it is reestablished in the daughter nuclei. The restoration of TADs can be explained by cohesin loading and loop extrusion, since the CTCF binding sites that demarcate most TAD boundaries are encoded in the DNA. Similarly, the reestablishment of compartments can be determined by histone marks that are inherited after mitosis (Alabert et al. 2014; Festuccia, Dubois, et al. 2016; Teves, An, Anders S Hansen, et al. 2016). However the question remains how chromatin organization is established in the first place during early development. Recent single-cell (or low-input) Hi-C studies in mouse oocytes and zygotes after fertilization provided some initial insights into this process (Du et al. 2017; Flyamer et al. 2017; Gassler et al. 2017). Interestingly, Hi-C data show a marked compartmentalization in mouse sperm but not in oocytes. This striking difference is attributed to (a) differences in compaction (a tightly compacted paternal genome versus a more decondensed maternal genome) and (b) differences in transcription during G1 phase. However, after fertilization, the difference is gradually decreased, and after eight cell divisions, both maternal and paternal genomes acquire fully fledged A and B compartments. Future work may use single-cell transcriptomics approaches to further understand the links between the observed chromatin reorganization and transcription in early developmental stages.

Conclusion. In this review, we have tried to summarize current knowledge about the 4D organization of the chromatin fiber, our understanding of its mechanisms, and some of the functional consequences of this organization, while also pointing to open questions and future research. Although necessarily incomplete, we hope that this review will help stimulate further work in this highly dynamic field at the crossroads of genetics, cell biology, and physics.

Future issues.

1. Development of a complete polymorphic model of chromatin fiber from the scale of nucleosomes to entire chromosomes would further enhance our understanding of chromatin folding at various scales.

2. Visualizing loop extrusion by cohesin in live cells will be important to test current models of TAD (topologically associated domain) formation.

3. New methods are needed to understand the effect of TADs on enhancer–promoter interactions and in the initiation or potentiation of transcription.

4. The molecular mechanisms underlying the segregation of A/B compartments require further investigation.

5. Formation of chromosome territories and nuclear bodies in the context of compartments, TADs, and phase separation needs to be clarified.

Glossary.

ChIP-Seq: an immunoprecipitation technique that provides the genomic locations where a protein of interest binds.

Chromosome territories: largely non-overlapping nuclear substructures first identified by FISH experiments using probes that specifically cover entire chromosomes.

Cohesin: a ring-like complex previously known for its role in holding the two sister chromatids together after replication.

Contour length: the distance between two points on a polymer as measured when walking along the chain, which can be much larger than the Euclidian distance between these points.

Compartments: megabase-scale regions of the genome, originally evidenced by checkerboard patterns in Hi-C maps, representing domains of increased interactions.

Enhancer: a non-coding region of DNA that can exert regulatory control on nearby or distant gene promoters by bringing transcription factors in contact with the transcription pre-initiation machinery.

Equilibrium (thermodynamic): the state of lowest free energy, where there is no net force acting on the system and statistical quantities such as temperature or pressure remain constant.

Euchromatin: initially identified by electron microscopy as less electron-dense chromatin regions in the nucleus, it contains most active genes.

Heterochromatin: initially identified by electron microscopy as electron-dense chromatin regions, it is subdivided into constitutive heterochromatin (never transcribed) and facultative heterochromatin (reversibly silenced genes).

Hi-C: a genomic technique that maps the contact frequency of any two DNA fragments sufficiently close in space to be captured by cross-linking.

Histone modification: the addition of small residues (e.g., acetyl, methyl, or phosphate groups) to different histones at different amino acid positions.

LAD: a megabase-scale chromatin domain shown to physically interact with the nuclear lamina by DamID.

Liquid–liquid phase separation: a thermodynamic process where two immiscible liquids progressively segregate into two phases, like oil and water (also termed demixing or coacervation).

Loop: two distant loci of the same chromosome maintained in close proximity, at least transiently, by other molecules; identified as peaks in Hi-C maps.

Nucleolus: the largest membrane-less organelle of the nucleus; it contains a low level of DNA and a high level of ribosomal RNA.

Persistence length: a measure of the rigidity of a polymer; it can be defined geometrically as the length over which tangent vectors remain correlated to each other.

Stripe/flame: a line of high contact frequency along a TAD border, arising when a locus is in contact with the entire TAD region.

TAD: a chromatin region of 40 kb to 3 Mb in size appearing along the diagonal of Hi-C maps as regions of enriched intrachromosomal contacts.

Chapter 3

Transcription in the nucleus

After a description of the content of the nucleus (section I.2), we now turn to a review of recent knowledge about transcription in mammalian cells and its regulation. We emphasize how the spatial constraints presented above and the complexity of the protein-protein interactions taking place in the medium influence the diffusion of transcription factors, and thus transcription itself.

More specifically, we review in section 3.1 some of our understanding of transcription in mammalian cells, and how chromatin and the arrangement of genomic elements (section 3.1.1) interplay with the complex core transcriptional machinery (section 3.1.2). In section 3.2, we review key concepts involved in the regulation of transcription, including cooperativity among transcription factors (TF; section 3.2.1), histone modifications, changes in the nuclear environment, such as chromatin density or mobility (section 3.2.2), phase separation (section 3.2.3) and enhancer-promoter contacts (section 3.2.4). In the last section, (section 3.3), we put a specific emphasis on the diffusion of transcription factors. After a brief theoretical introduction (section 3.3.1), we highlight some of the main findings that were provided by the study of TF diffusion (section 3.3.2).

3.1 Overview of transcription in mammalian cells

Transcription is the process by which a gene is being transcribed in a messenger RNA. Despite the fact that in *in vitro*, viral and bacterial systems, very few protein complexes are needed for RNA production, mammalian systems have evolved a considerable amount of regulations that are organized at various levels, from regulatory genetic sequences to a complex cocktail of co-activators called transcription factors.

In this section, we highlight some basic concepts about transcription in mammalian cells. Another introduction can be found in (Levine, Cattoglio, and Tjian 2014). We aim at detailing the basic concepts needed to motivate our study of the dynamics of the transcription factors c-Myc, as highlighted in section II.3.

3.1.1 Genomic elements

We first start by a description of genomic elements involved in transcription. In mammalian cells, it is usually acknowledged that to activate a gene, it must be located downstream of an active promoter sequence, bound by the RNA polymerase. Furthermore, additional factors and enhancer-promoter contacts are required to transcribe a gene.

1. **The promoter.** The first key genomic element to activate a mammalian gene is a promoter sequence. It is located upstream of the gene body. The *core promoter*, a $\sim$ 80 bp sequence, is the place where the pre-initiation complex (PIC) assembles (detailed in 3.1.2). The extended promoter sequence is usually bound by co-activator proteins and other TFs that can enhance or repress transcription.

The structure and function of the core promoter has been dissected in details, and many motifs involved in the recruitment of some PIC subunits have been identified (see James T. Kadonaga 2012 for a review). For instance, the most widely known TATA-box recruits the TATA-box binding protein (TBP), a TFIID subunit. The BREu and BREd elements are bound by TFIIB, the initiator sequence (Inr) is bound by TAF1 and TAF2, the downstream promoter element (DPE) is bound by TAF6 and TAF9 (Tamar Juven-Gershon and James T. Kadonaga 2010). Not all these elements are present in all core promoters. For instance, only 10-20% of the human promoters contain a TATA-box element (Tamar Juven-Gershon, Cheng, and James T. Kadonaga 2006). Finally, from a functional perspective, the presence/absence of some elements has been shown to influence the regulation of different transcriptional regulation programs (T. Juven-Gershon, Hsu, and J. T. Kadonaga 2008; Tamar Juven-Gershon and James T. Kadonaga 2010; Weingarten-Gabbay, R. Nir, et al. 2017, reviewed in Weingarten-Gabbay and Segal 2014).

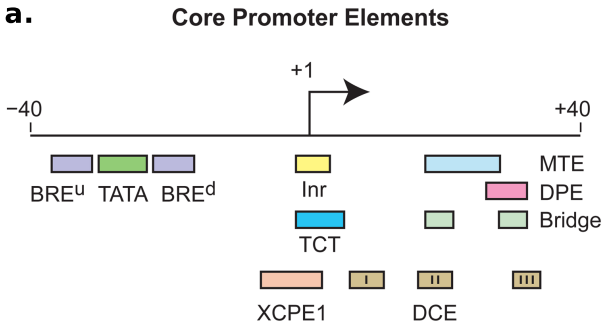


Figure 3.1: **a.** Some core promoter elements for transcription by RNA polymerase II. The locations of the motifs are drawn roughly to scale. The BREu, TATA, Inr, MTE, DPE, and TCT motifs have been found in both *Drosophila* and humans. These motifs are typically found in focused core promoters, although there are probably Inr-like elements in dispersed promoters. There are no universal core promoter elements that are found in all promoters. Moreover, it is likely that many other core promoter motifs remain to be discovered. The functional properties of a core promoter are determined by the presence or absence of specific core promoter motifs. For example, some enhancers will activate transcription from DPE-dependent core promoters but not from TATA-dependent core promoters. (source: James T. Kadonaga 2012).

2. **Enhancers.** In mammalian cells, most genes display reduced activity in the absence of distal regulatory sequences called *enhancers* (Leemans et al. 2018). Enhancers are short genomic sequences (50-1000 bp) that display a regulatory potential. Although difficult to precisely quantify, most of the distal enhancer sequences identified so far are located within 500 kb or a gene (median distance of 125 kb; van Arensbergen, van Steensel, and Bussemaker 2014), but some of them have been found as far as 1.7 Mb downstream of their target genes (von Paleske et al. 2014) and more than half of them can skip one or two genes to activate their target gene (van Arensbergen, van Steensel, and Bussemaker 2014). In some cases, authors have argued that megabase-size genomic sequences carry an enhancer-type property, leading to the concept of "regulatory archipelago", in which many different, weak enhancers are scattered over long genomic distances (Montavon et al. 2011).

   The identification of enhancer sequences has been a long-lasting quest, and so has been the identification of the target genes of a given promoter. The number of enhancers sequences in cells could be as high as 1 million, giving an average of 50 enhancers per gene.

   Some transcription factors (TFs) are thought to be the proteins mediating the interaction between enhancers and promoters. TFs bind to a specific sequence, termed the binding *motif*, through their DNA-binding domain.

3. **Transcription factors binding motifs.** TF binding motifs, or TF binding sites (TFBS) are short (6-20 bp) DNA sequences specifically recognized by TFs. Some TFs can bind a family of motifs, or degenerate motifs. The role of this low specificity has been demonstrated as an evolutionary strategy to regulate with one factor (one Hox TF) many genes (Crocker et al. 2015). It is thought that these motifs appeared in the genome through random point mutations, but also through radiation of transposable elements that contained clusters of TFBS. This latter mechanism could explain the complex transcriptional co-regulation of mammalian cells.

   Unlike in bacteria that have a short genome and TFs that recognize long motifs (16-20 bp motifs), mammals have larger genomes and their TFs usually recognize short motifs (6-8bp). Wunderlich and Leonid A. Mirny 2009 estimated the information contained in these motifs, and found that a 6-bp sequence was not sufficient to precisely specify a given binding region in such a big genome (to target a TF to a unique location in the genome), contrary to the bacterial case (a long motif in a small genome). In this setting the only way of getting single-locus specificity is to enable some cooperativity between TFs.

   Considerable efforts have been dedicated to the identification of the binding motif of DNA-binding proteins, and many approaches have been implemented, ranging from *in vitro* assays to bioinformatics based on ChIP-seq or massively parallel reporter assays (Kinney, Tkačik, and Callan 2007; Barnes et al. 2018)

   Having a TFBS at a given locus doesn't mean that a TF will be found there, and the link between the presence of a TFBS and the binding of the TF is not fully understood. It at least involves chromatin accessibility and the presence of cofactors (Jian Yan et al. 2013).

3.1.2 The core pre-initiation complex (PIC) machinery

To transcribe a gene, the RNA polymerase II (RNA Pol II) must be loaded on its promoter. This operation involves the step-wise assembly of a multi-protein (dozens of proteins), megadalton size complex called the transcription pre-initiation complex (PIC).

The assembly of the PIC is regulated by many external factors that can facilitate or repress transcription. Several series of techniques have greatly advanced our understanding of PIC assembly. The first ones are biochemical purification, cryo-electron microscopy (cryo-EM) and crystallography. The second ones are live imaging. Both draw a complementary picture of transcription.

1. **Static picture.** Transcription is usually divided in several, independently regulated steps: initiation, pause, elongation, termination. Most of the regulation studies focused on initiation, pause and elongation.

   (a) *Step-wise assembly of the PIC.* A combination of biochemical studies, cryo-EM and other biochemical characterizations, including the use of a collection of transcription inhibitors (Bensaude 2011) allowed to draw a picture of the step-wise assembly of the PIC (reviewed in Sainsbury, Bernecky, and Cramer 2015).

   The PIC is subdivided in several complexes: TFIIA, TFIIB, TFIID, TFIIF, TFIIH, each of them composed of several proteins. For most of them, the exact structure has been at least partially resolved (for instance Murakami et al. 2015; Louder et al. 2016). A canonical vision of PIC assembly is as follows (Nikolov and S. K. Burley 1997; Sainsbury, Bernecky, and Cramer 2015): (1) the core promoter sequence is recognized by the TATA-box binding protein (TBP; a TFIID subnit, that also contains $\sim$ 15 TBP-associated factors, or TAFs), (2) the subsequent recruitment of TFIIA and TFIIB in an intermediate complex, (3) the recruitment of PolII and TFIIF to form the core PIC, (4) the recruitment of the regulatory subunits TFIIE and TFIIH, (5) the creation of a transcription bubble and the unwinding of DNA and (6) the onset of elongation and the disassembly of the PIC leaving PolII alone

along the gene body, (7) promoter-proximal pausing and finally (8) the elongation step (Figure 3.2).

Interestingly, these transitions were first biochemically evidenced before being imaged with cryo-EM and crystallography. This was possible by the use of various transcription inhibitors (Bensaude 2011) that allowed to lock transcription in a given state. For instance, DRB and flavopiridol acts on CDK9 and blocks the transition from initiation to elongation.

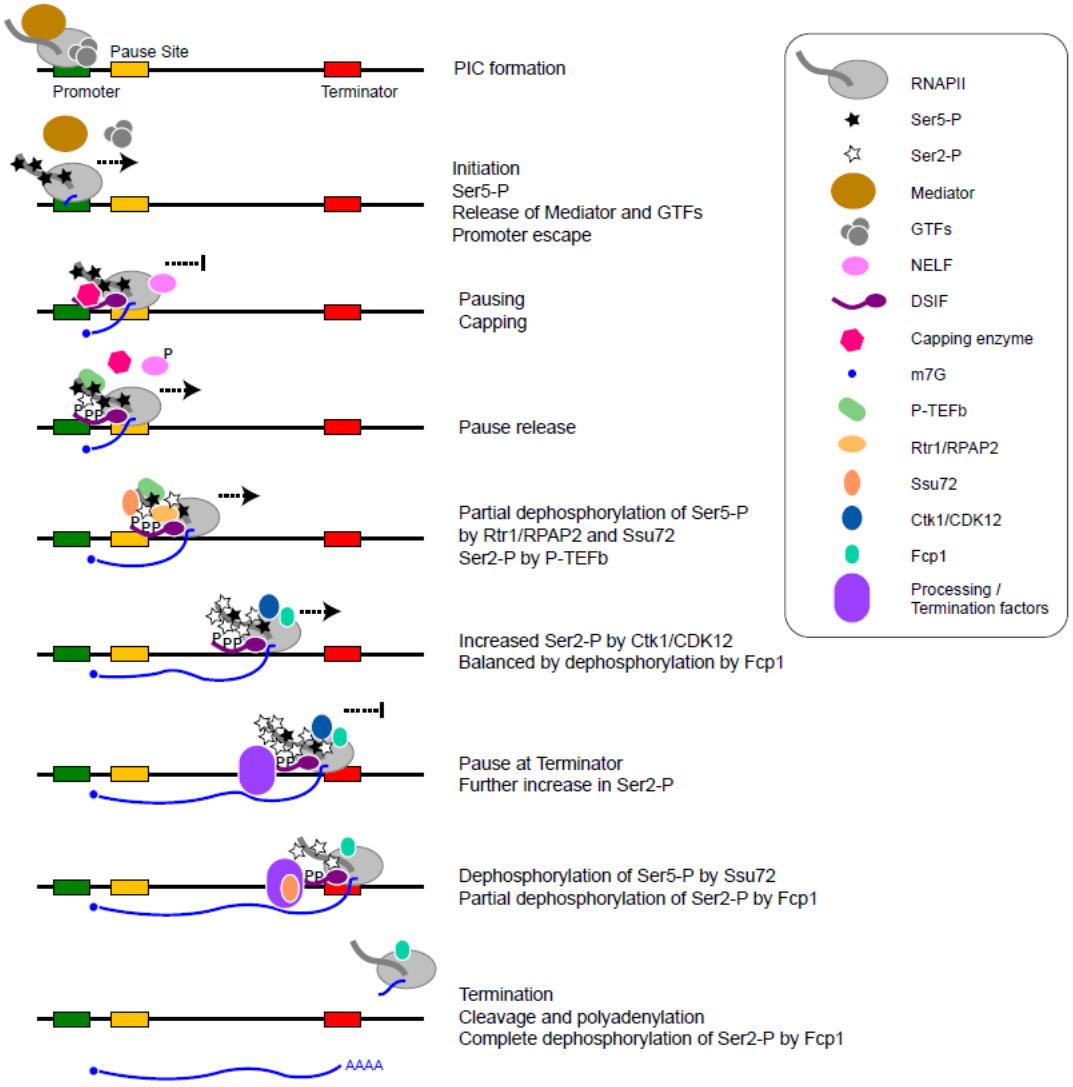


Figure 3.2: **Overview of mammalian transcription at a single locus.** Step-wise progression of transcription from PIC assembly to transcription termination. The main steps of transcription are presented, together with the main proteins and complexes involves (source: Jeronimo, Collin, and Robert 2016).

(b) *Pol II C-terminal domain (CTD).* Many of these transitions have also be linked to a change in the phosphorylation of the C-terminal domain (CTD) of the Rpb1 subunit of RNA PolII. PolII CTD is a highly repetitive sequence, that contains 52 repeats of seven aminoacids: YSPTSPS, many of them acting as phosphosites. Crucially, most of these 52 repeats undergo dramatic changes in phosphorylation that are coupled to the transitions between the various states of transcription, as detailed above (and in Figure 3.2), (Palancade and Bensaude 2003).

In particular, initiating polymerase is characterized by a CTD phosphorylated on Ser5 (on the 5th serine of the repeats), and the transition to elongation is characterized by the removal of Ser5 and the addition of Ser2 phosphorylation by P-TEFb. Getting to know which repeat of the CTD is preferentially phosphorylated has been extremely challenging, but several studies indicate that not all repeats are functionally equivalent. Studies are complicated by the high number of repeats that prevent traditional mass-spectrometry analysis. Recent genome-editing techniques might overcome this limitation by adding identifiable spacers in the PolII-CTD sequence (Schüller and Eick 2016).

(c) *Around the PIC.* Most of the transcription regulation (briefly reviewed in section 3.2) occurs at the transition between these steps. Indeed, many transcription factors and co-regulators transiently interact with the PIC. For instance, in many organisms including yeast and human, the SAGA (Spt-Ada-Gcn5 acetyl-transferase; Warfield et al. 2017; Baptista et al. 2017) and the Mediator (Kagey et al. 2010; Koutelou, Hirsch, and Dent 2010) complexes are involved in transcription initiation. Some evidence suggest that the Mediator complex might be involved in enhancer-promoter looping.

Similarly, many other co-activators and cofactors are involved in transcription regulation. Key questions that usually arise are (1) what is their precise role in the sequence of events leading to transcription elongation? (2) how are they recruited to the PIC? (3) what is their dynamics of recruitment. We briefly review some of the knowledge of the PIC assembly in the next section.

2. **Dynamics.** Most of the findings presented in the previous section are derived from biochemical analysis and provide little information about the dynamics and timing of transcription. In this section, we present how imaging approaches provided more precise information about how transcription happens in time.

(a) *Global picture of the dynamics of transcription initiation.* A first understanding (Darzacq et al. 2007) of the dynamics of transcription was provided by a FRAP study that combined transcription on an artificial gene array inserted at one locus of U2OS cells (Janicki et al. 2004) and transcription inhibitors (Bensaude 2011) through kinetic modeling.

The authors evidenced that transcription initiation is a highly inefficient process, with more than 99% of assembled polymerases not giving a full-length messenger RNA, with a polymerase spending $\sim$ 1 min in the initiation state. These results were later confirmed using an endogenous Rpb1-GFP knock-in (Steurer et al. 2018; Price 2018), and the steps around promoter-proximal pausing were further dissected.

Furthermore, the residence time of TBP on a given promoter has been quantified, and is on the order of minutes (Zaidi, Auble, and Bekiranov 2017; Teves, An, Bhargava-Shah, et al. 2018)

(b) *Getting up to speed: measuring PolII speed.* Once PolII has escaped from the promoter, it usually undergoes a promoter-proximal pause followed by the full elongation (Jonkers and Lis 2015). Using a knock-in and a different set of inhibitors, Steurer et al. 2018 gained more insights on how PolII pauses and elongates. First, they found that in cells, PolII spends $\sim$ 40 s in the paused state. Second, this study found that PolII elongates at a rate around 2 kb/min, where Darzacq et al. 2007 estimated a rate of 4.3 kb/min. More generally, the elongation speed of PolII has been estimated using several techniques, yielding usually consistent results (see for instance D. R. Larson et al. 2011, that estimates an elongation speed between 1.2 and 2.7 kb/min, depending on the gene, or Garcia et al. 2013 that estimate a rate of 1.5 kb/min in *Drosophila*).

Finally, the dynamics of PolII on a gene has also been studied *in vitro* at the single-bp level using optical trap (Righini et al. 2018). Such an approach provides complementary biophysical information about PolII dynamics.

(c) *Producing multiple mRNAs: gene bursting.* Taking a step back, the number of polymerases at a given promoter and elongating on a gene is also regulated. First, the transcription of many genes occurs in *bursts*: several mRNAs are produced during short period, interspersed by long periods of gene inactivity. The factors that influence bursting have not been fully elucidated, but probably include the dynamics of TBP turnover (Werven et al. 2009; Sprouse et al. 2008) accumulation of torsional stress on DNA, the presence of cofactors such as the Mediator complex (Tantale et al. 2016), the mode of exploration of TFs (compact vs. non-compact; Meyer et al. 2012) and PolII pausing dynamics at this gene (W. Shao and Zeitlinger 2017).

3.1.3 Histones on the path of transcription

The vision of transcription presented in the previous section assumes that DNA behaves nicely as a substrate for transcription. However, as detailed in section I.2, DNA is wrapped with high affinity around nucleosomes, a ~ 10 nm diameter structure that can block (or at least slow down) transcription. Furthermore, histones contain low-complexity tails (N-terminal domains; NTD) that can be heavily post-translationally modified, and whose modifications are strongly correlated with transcription.

1. **Getting through nucleosomes.** First of all, nucleosomes act as a barrier to transcription, since a significant amount of energy needs to be spent to unwind DNA from a nucleosome such that a transcribing PolII can come across ($\sim 40 k_B T$; Mack et al. 2012). How mechanistically PolII gets across nucleosomes is not fully understood, but a recent cryo-EM studies proposed preliminary models (Kujirai et al. 2018). Furthermore, *in vitro* NMR and single-molecule studies demonstrated how chromatin remodelers greatly help the polymerase to get through nucleosomes (K. K. Sinha, Gross, and Narlikar 2017; Fitz et al. 2016).

   Finally, because promoters are occupied a significant fraction of the time by inefficiently initiating polymerases, a "phasing" of nucleosomes has been shown to arise from the stochasticity of nucleosome remodeling: nucleosomes appear regularly spaced around the promoter (Parmar, Marko, and Padinhateeri 2014; Chereji and Clark 2018). For long, it has been thought that nucleosomes had a more regular spacing in active regions compared to inactive region, but a recent study using nanopore sequencing suggests that inactive regions might actually exhibit stronger phasing than the active ones (Baldi et al. 2018).

2. **Role of histone modifications.** Histone tails can be post-translationally modified, and these histone modifications constitute transcriptionally-relevant landmarks. A complex collection of factors is involved in modifying and de-modifying histones, and these factors are under tight regulation. The collection of existing histone modifications is enormous (Figure 3.3; reviewed in Bannister and Kouzarides 2011), and new histone modifications are regularly discovered (Lawrence, Daujat, and R. Schneider 2016). Many of the PIC components display a kinase (or other) activity and can target or nucleosomes directly or recruit chromatin modifiers.

   (a) *At the genome-wide level.* Histone modifications are not evenly spaced on the genome, and immunofluorescence microscopy of mitotic chromosomes showed large clusters of regions preferentially modified histones (Figure 3.3b; Terrenoire et al. 2010), this discovery gives a foundation to the genomic compartments phase-separation models presented in section 2.5.

   (b) *At the genomic element level.* It has been shown that many genomic loci can be characterized by a specific pattern of histone modifications. For instance, active genes are usually associated with H3K4me3 (trimethylation of lysine 4 of histone 3) and H3K27ac. Conversely, inactive genes usually display a low level of these modifications and a high level of H3K27me3. Furthermore, enhancer sequences will also be characterized by a specific histone pattern: high level of H3K4me1 and medium level of H3K4me3. As such, whole genomic partitions can be derived almost solely based on the sequencing of histone modifications.

(c) *At the single-gene level.* Around a prototypical active gene body, histone modifications are specifically positioned, and some of them are shown in Figure 3.3c. The promoter of active genes is enriched in H3K4me3, while the levels of H3K36me increases across the gene body. The mechanisms by which this pattern arises is unclear, but the co-localization of histone modifiers might explain part of this distribution (Morris et al. 2013).

(d) *Histone modifications and transcription: cause or consequence?* Despite a very clear association between histone modifications and transcription (histone modifications are predictive for gene expression; Karlic et al. 2010; Pradeepa et al. 2016), the exact link between them is not clear, and whether histone modifications are a cause or a consequence of transcription is still an open question (Figure 3.3d and a review in Howe et al. 2016).

Several conflicting lines of arguments exist: a first one looks at the effect of the removal of H3K4me3, one of the main marker of active genes, and find little effect on transcription. Second, *in vitro* transcription of a chromatinized sequence using the mammalian transcription machinery does not require H3K4me3 (both reviewed in Howe et al. 2016).

Another line looked at the impact of histone acetylation on transcription. A first study (Stasevich et al. 2014) looked at the level of histone acetylation on induced gene expression (during an induced glucocorticoid response) and found that arrays with initially more histone acetylation had higher levels of PolII binding after the response. Another study looked at burst frequency (Nicolas et al. 2018) and showed using live imaging that perturbations in acetylation influenced mRNA production. Finally, on the enhancer side, it has been shown that occupancy by a given set of TFs is a more accurate predictor of active enhancer than histone modifications or chromatin accessibility (Dogan et al. 2015).

All in all, more studies are needed to clearly disentangle the role of histone modifications in transcription.

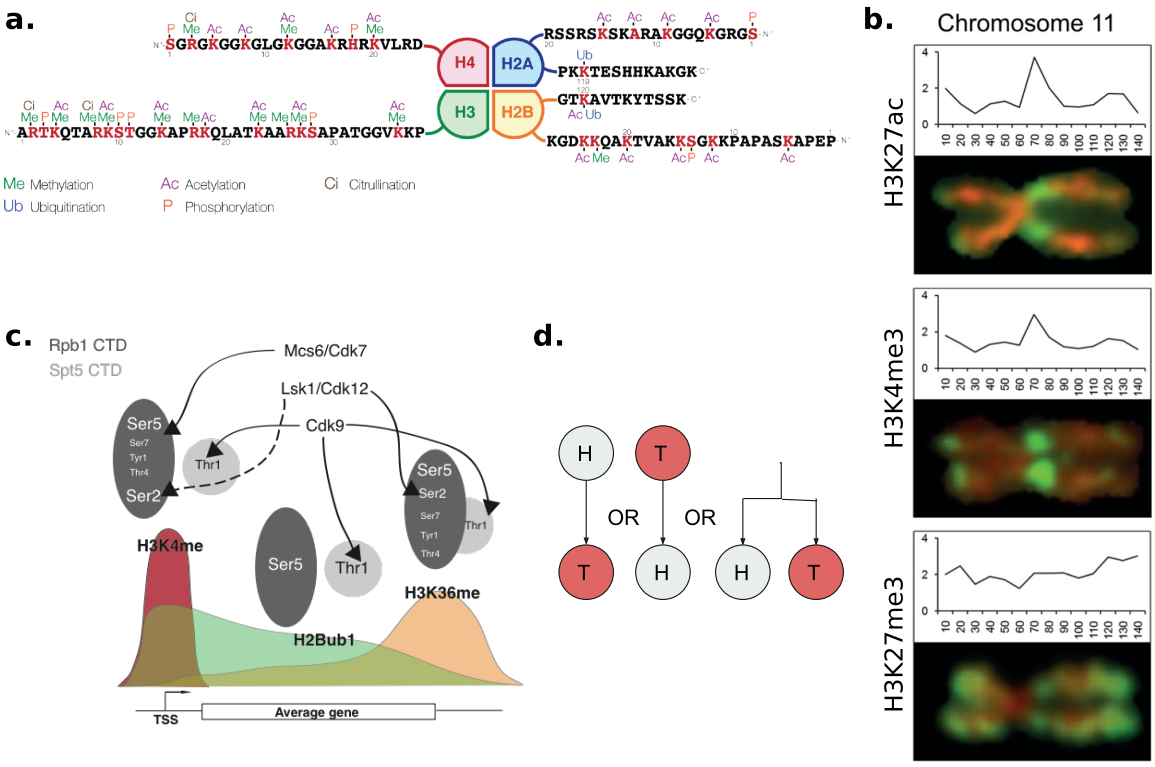


Figure 3.3: **a.** Histone tails can be post-translationally modified (source) **b.** Global pattern of histone modifications at the scale of a chromosome (from Terrenoire et al. 2010) **c.** Pattern of histone modifications around a typical mammalian genes, and some chromatin modifiers responsible for these modifications. co-transcriptional histone modifications by the Rpb1 and Spt5 CTDs. Arrows connect the transcription-associated CDKs to their targets on the Rpb1 or Spt5 CTDs above the three major co-transcriptional histone modifications (H3K4me-histone H3 lysine 4 methylation, H2Bub1-histone H2B ubiquitylation, H3K36me-histone H3 lysine 36 methylation). Solid arrows indicate positive regulation through the indicated phosphorylation event, whereas dotted arrows indicate de-phosphorylation event (source: Tanny Lab, McGill University) **d.** Potential causal relationship between gene expression and histone modifications (borrowed from a joint presentation with J. J. Parmar).

3.1.4 Transcription factors, pioneer factors & mitotic bookmarking

The previous sections emphasized the role of genomic elements (section 3.1.1) and of the PIC (section 3.1.2). We now turn our attention to the proteins involved in transcription, and in particular to transcription factors. Transcription factors are proteins involved in transcription regulation. They are usually subdivided between *general transcription factors* (GTFs; that are needed at the promoter of all active genes and *regulatory transcription factor* (that are involved in the specific regulation of some genes), despite these categories being loosely defined in practice. The components of the PIC are GTFs.

Many other TFs are involved in the regulation of transcription. We briefly highlight here two of their properties. They are usually proteins constituted of at least two domains: a DNA-binding domain that contains a motif recognition sequence, and a trans-activation domain (TAD), that is responsible for gene activation. Some factors are particularly important in transcription since they can bind to nucleosomes, and are thought to be able to activate transcription from a silent locus.

1. **Trans-activation domains.** The properties of DNA-binding domains are understood in more and more details. Indeed, these domains are usually well structured in complex with DNA, and

can be crystallized. As such, we now have the crystal structure of a high diversity of DNA binding domains.

Conversely, trans-activation domains of TFs are usually low complexity sequences, disordered and can contain tens of repeats of the same amino-acid sequences: the traditional lock-and-key mechanism at play to explain DNA specificity does not work for unstructured domains and they cannot be easily crystallized. Most (82-94%) mammalian TFs have a disordered domain (Jiangang Liu et al. 2006), and most of them (73-94%) have a disordered TAD. How these domains achieve trans-activation is not clear.

Several properties of TADs have been proposed to explain their function, and many experiments focused on the charge pattern of the TAD. For instance, the charge of the domain has been shown to influence TF properties (Y. Gao et al. 2017). Also, the patterning of charges has been shown to be a major driver of trans-activation properties (Sherry et al. 2017). Finally, the number of disordered repeats has also been studied (Gemayel et al. 2015). On a more systematic analysis (Marsh and Forman-Kay 2010) measured the compaction of 32 proteins and linked it with sequence properties.

2. **Pioneer factors.** A second property of some transcription factors is their ability to bind nucleosomes. This property is thought to be key to gene reactivation (Zaret et al. 2008). Indeed, inactive genes are thought to be entirely nucleosomal, and their promoter is usually not accessible. Thus, a factor able to bind nucleosomes might be able to recruit chromatin remodelers and to progressively remove the histones from the promoter, allowing a gene reactivation. This hypothesis was investigated in (Soufi et al. 2015) and evidenced that Oct4, Sox2 and Klf4 have the ability to bind nucleosomes. Over the time, some other factors such as FoxA2 and GATA4 have been shown to exhibit pioneering properties (Donaghey et al. 2018). More generally, there is a complex interplay between genome organization and TF activation, and (Stadhouders et al. 2018) found that TFs drive genome reorganization at multiple architectural levels during reprogramming.

3. **Bookmarking.** While pioneer activity allows to reactivate a silent gene, another key question is how to know which genes were active before mitosis, to reactivate only these ones. In other terms, how do cells bookmark gene activity? This question has lead to several hypothesis. One of them proposed that some general or specific transcription factors remain bound to chromatin during mitosis. This hypothesis is consistent with the fact that the promoter of active genes remains accessible in mitosis (Teves, An, Anders S Hansen, et al. 2016) and that TBP and Esrrb remain bound to their promoters (Festuccia, Dubois, et al. 2016; Teves, An, Bhargava-Shah, et al. 2018). As such, TBP has been proposed to be one bookmarking factor, since it can remain associated with DNA for the duration of mitosis. Conversely, other TFs such as Oct4 and Sox2 clearly do not act as bookmarking factors (Teves, An, Anders S Hansen, et al. 2016; Festuccia, Owens, et al. 2018).

3.2 Mechanisms of transcription regulation

In section 3.1, we presented the main categories of factors involved in transcription. This initial picture explains how a gene can be transcribed, but it does not explain why a given gene can be transcribed and not another. In this section we review some mechanisms involved in transcriptional regulation. These mechanisms can be classified in four categories: (1) the cooperative effect of TFs (section 3.2.1), (2) changes in the chromatin environment (section 3.2.2), (3) changes in local concentrations and phase separation (section 3.2.3) and (4) regulation of enhancer-promoter contacts (section 3.2.4).

3.2.1 Combinatorics of the presence of TFs

1. **Presence/absence of a single TF.** A first level of regulation takes place by the regulated expression of TFs. Indeed, the expression of some transcriptional program is under the control of

some TFs that are only expressed in a given cell type. This is for instance the case for many TFs involved in development, such as GATA4, a factor specifically expressed in the heart and that regulates cardiac development (RNA profile). More interestingly, some components of the PIC (TAFs), such as TAF9B are differentially expressed and regulate different groups of genes (despite being usually termed as GTFs). For instance, TAF9B is specifically induced and required for motor neuron differentiation (Herrera et al. 2014).

2. **TFs cooperate to activate a gene.** Second, many studies show that a combination of factors (or the multiple binding of one factor) is required to activate a gene (for a theoretical justification, see Wunderlich and Leonid A. Mirny 2009).

   (a) *Several copies of the same TF.* In some cases, gene activation requires the binding of several copies of the same protein. For instance, it has been proposed that multiple copies of P-TEFb are needed in order to release a paused polymerase (X. Lu et al. 2016).

   (b) *Cooperation between several types of TFs.* The recruitment of some GTFs is mediated by the binding of some TFs. For instance the oncogene p53 mediates the recruitment of the GTF TFIID (Coleman et al. 2017). In most of the cases, the cooperation between TFs involves TFs from different families, a theory that matches the genome-wide binding pattern of TFs, that tend to bind in clusters. For instance, Oct4, Sox2 and Klf4 tend to bind the same enhancer elements, and it is known that Oct4 and Sox2 bind cooperatively (Chronis et al. 2017). More generally, classes of TFs that bind cooperatively can be defined based on their ChIP-seq patterns (Beck et al. 2014 and Alexander Stark, personal communication).

   Cooperativity can also be evidenced by the fact that not only the DNA-binding domain is important in directing a TF to a specific locus but also the trans-activation domain. Indeed, a change in trans-activation domain can dramatically modify the DNA-binding pattern of a TF, as evidenced in (W. F. Lim et al. 2016).

   Mechanistically, cooperativity can either be mediated through direct protein-protein interactions, but also be indirect. A biophysical model suggest that nucleosomes can also mediate TF cooperativity without any direct contact (Leonid A. Mirny 2010).

3.2.2 Chromatin environment

A second level of regulation takes place at the level of the chromatin environment, in which parameters such as the position with respect the nuclear periphery, the local density of chromatin and the fluctuations of chromatin have been occasionally shown to influence gene expression. The mechanisms of this regulation, however, are still unclear.

1. **Chromatin environment and gene (re)-positioning.** Chromatin environment might influence gene regulation in several ways.

   (a) *Lamina-Associated Domains (LADs) constitute a repressive environment.* First, a gene located in a lamina-associated domain is usually repressed compared to the activity it would have outside the LAD (Guelen et al. 2008; Akhtar et al. 2013; Leemans et al. 2018 and van Steensel and Belmont 2017 for a review). Why LADs constitute a specific repressive environment is not specifically understood.

   (b) *Gene reposition to the center of the nucleus when they activate.* Second, it has been shown that some genes relocate to the center of the nucleus when they activate (Tumbar and Andrew S. Belmont 2001). The mechanism of such process is not known, but nuclear F-actin might be involved, as it is for the re-localization of DNA breaks (Caridi et al. 2018), or other, still unexplored processes might be at play (J. Kim et al. 2018).

2. **Chromatin density.** For long, chromatin density itself has been proposed as a way of explaining why some regions are active and other inactive. In the initial theory, regions of high chromatin density would be less accessible and thus not permissive for transcription. This idea, however, can be discussed in several ways.

   (a) *Can dense regions prevent transcription?* First, in terms of accessibility, dense heterochromatin and less compacted euchromatin are often similarly accessible to a diffusing transcription factor (see section 3.3.1). On the other hand, ChromEMT images produced by the O'Shea lab (Ou et al. 2017) showing nucleus slices at the single-nucleosome level revealed that in dense regions, the PIC might not have the space to assemble, despite all the factors being able to diffuse (Clodagh O'Shea, personal communication). Furthermore, it has been suggested that decondensation is sufficient to induce gene expression (Pierre Therizols et al. 2014; Benabdallah et al. 2017).

   (b) *From density to accessibility.* This idea that dense regions are less permissive to transcription got further complicated by the fact that "chromatin accessibility" measurements were performed using several enzymes followed by sequencing (DNAse hypersensitivity assay and ATAC-seq for instance). It is not exactly clear what these assays measure, and whether their output is also a proxy for chromatin density is still unclear.

3. **Chromatin dynamics.** Finally, it has been proposed that chromatin fluctuations, rather than chromatin density, might explain gene activation. Indeed, at very small scale a slow-moving chromatin might hamper the assembly of the PIC, while a fast-moving chromatin might sample a sufficient number of conformations to allow the scaffold to build, even in high-density environments. This idea was proposed based on data on harshly treated cells followed by fluorescence lifetime imaging (FLIM; Auduge et al. 2018). This idea is reinforced by the fact that NMR studies have shown that the presence of the linker histone H1 involved in heterochromatin reduces the fluctuations of chromatin (Stützer et al. 2016). Also, the rigidity and dynamics of chromatin seem to evolve as transcription progresses (Germier et al. 2017).

3.2.3 Phase-separation & local concentrations

A third level of regulation occurs when concentrations of factors are locally increased, which can lead to locally increased reaction rates. More and more, this concept is mixed with the concept of phase separation, in which phases made of different components separate because they share different chemical properties.

1. **Genes local enrichment** From a genomic perspective, co-regulated genes tend to cluster. They tend to be located in the same TAD, or to form large clusters in Hi-C or derived techniques. This was initially evidenced using 3C (Osborne et al. 2004; Schoenfelder et al. 2010). This way, co-regulated genes might share some limiting resources such as PolII, whose high local concentrations is required to overcome its inefficient initiation.

2. **PolII local enrichments.** In addition to heterogeneities in RNA and DNA concentrations, (reviewed in section 2.1.1), PolII displays strong and dynamic local enrichments.

   Indeed, live PALM imaging suggested that short-lived clusters (a few seconds) existed in the cells (Cisse et al. 2013; Rickman and Bickmore 2013). Despite some initial criticisms (Z. W. Zhao et al. 2014), this finding was confirmed and it was shown that these clusters co-localized with active genes on a selected example (Cho, Jayanth, Mullen, et al. 2016) and that a link could be found between the size of the cluster (ie. the increase in local concentration) and gene expression (Cho, Jayanth, Brian P. English, et al. 2016).

3. **Pol II phase-separation.** Going beyond local enrichment, it was suggested that PolII can undergo reversible phase-separation, and that this phase-separation is regulated by phosphorylation (Kwon et al. 2013). Furthermore, this phase-separation behaviour was linked to the CTD of Rpb1, and *in vivo* modifications of the length of the CTD lead to changes in the size of the Rpb1 clusters (Boehning, Dugast-Darzacq, Rankovic, Anders S. Hansen, T.-K. Yu, et al. 2018).

4. **Transcriptional phase-separation.** Finally, it has been proposed more recently that transcription might occur through a phase-separation process, in which the whole PIC scaffold, enhancers and co-activators would undergo a phase-transition. This was theoretically proposed in (Hnisz et al. 2017). Several lines of evidence tend to suggest that such a phenomenon might be at play. First, many TFs can undergo *in vitro* reversible phase separation (Boija et al. 2018), and thus they can contribute to such transcription aggregate. Second, such phase-separated aggregates seem to be also present *in vivo* (Sabari et al. 2018; Saey 2018). Third, it has been proposed that TF-induced phase separation might be sufficient to explain the presence of some TADs (Pereira et al. 2018). Fourth, in *Drosophila*, some genes appear to be simultaneously ON and OFF at the same time, suggesting that they might be residing in the same regulatory unit (B. Lim et al. 2018).

3.2.4 Enhancer-promoter contacts

Under the current dogma, (at least) one enhancer must contact a gene in order to produce mRNAs. This is thought to be mediated by enhancer-promoter looping, a phenomenon that has been widely investigated using 3C-based techniques. In this section we provide a critical review of the recently published literature that links enhancer and promoter contacts with gene expression, and show that an emerging body of experiments suggest that additional mechanisms compared to simple enhancer-promoter contacts might be at play. Previous perplexing questions were reviewed in (Andrew S Belmont 2014).

1. **How to transcribe a gene?** Let us first remind that outside mammalian systems, enhancers are neither present nor required to activate a gene. In bacteria, distal regulatory sequences are not needed for the RNAP to initiate transcription. Furthermore, an *in vitro* transcription reactions usually contains the following ingredients: (1) DNA template, (2) T7 RNA polymerase (from T7 phage, structure 1MSW), (3) nucleotides, (4) water. Thus a very minimal number of components are needed to transcribe a gene and the emergence of enhancer-promoters contacts is important for the regulation of this process, but not as a core requirement.

2. **Genome organization and enhancer-promoter contacts: evidence from 3C/Hi-C.** Let us then examine generous evidence from 3C and Hi-C. A long stream of papers focuses on "loops", that are regions of increased contact probability and that appears to link enhancer sequences with promoters and correlates with gene activity (S. S. Rao et al. 2014). Some of these enhancer-promoter contacts have been convincingly validated using multi-color FISH, where it was shown that indeed loci that appear in the same "loop" have a lower physical distance distribution than random loci located at the same genomic distance. This finding was reiterated when higher-resolution Hi-C and micro-C (Hsieh et al. 2016) were published (Bonev et al. 2017 and Stanley Hsieh, unpublished data). Thus, strong correlative evidence support the fact that enhancer-promoter contacts are associated with gene expression.

3. **Induction/disruption of enhancer-promoter contacts regulate gene expression.** Furthermore, in some limited cases, it was causally proven that disruption of an enhancer-promoter contact can result in transcriptional alteration and developmental defects. Indeed, (Lupiáñez et al. 2015) disrupted an enhancer-promoter contact by triggering a genomic inversion in the EPHA4 locus. This genomic inversion mimicked a genetic anomaly present in a patient, and could reproduce the same phenotype (a polydactyly disorder). Another example, detailed in

(Deng et al. 2014) in which the authors used an engineered zinc-finger protein to bring into contact an enhancer and a promoter in the beta/gamma-globin locus. This lead to the apparition of a "loop" in 3C and to the reactivation of a gene, further reinforcing the notion that when enhancer and promoters come into contact, gene activation occurs. A third example directly visualizes enhancer-promoter distances in Drosophila (Hongtao Chen et al. 2018) and finds that enhancer-promoter distances are lower when a reporter gene is active than inactive.

4. **Conflicting evidence about the function of enhancer-promoter proximity.** We now turn to a few recent papers that provide puzzling counter-evidence to the above-presented dogma.

   First of all, (van Arensbergen, FitzPatrick, et al. 2016) used a genome-wide sequencing approach to evaluate the ability of promoters to initiate transcription in the absence of any enhancer (in an episomal context). They found that there is a significant correlation between the expression level of the promoters in cells and without any enhancer, suggesting that promoters are the main determinant of gene expression, and not enhancer sequences.

   Second, in (Benabdallah et al. 2017), the authors focused on the *Shh* locus and observed an increase in the enhancer-promoter distance when the Shh gene is reactivated during development, where the traditional enhancer-promoter model would suggest a reduced distance. They progressively dissect the mechanism and find that gene expression is impaired when a "roadblock" is inserted between the enhancer and promoter. This experiment suggest that instead of a direct enhancer-promoter contact, a "message" might spread along the genomic DNA from the enhancer to the promoter. The authors also show that this mechanism relies on the poly-ADP-ribose polymerase.

   Finally, (Alexander et al. 2018) performed live cell imaging in mESC on the Sox2 locus in which the location of the enhancer and promoter were tracked, in addition to the Sox2 expression using a MS2 reporting system. Using a series of statistical analysis, the authors find that across the time scales they considered (up to several tens of minutes), there is no correlation between enhancer-promoter distance and gene expression.

5. **Towards a resolution?** Currently, the traditional enhancer-promoter contact model has a strong experimental support. However, it is not clear how accurate this model is in terms of dynamics. Can one imagine that a promoter remain "potentiated" several minutes after an enhancer-promoter contact? In that case, how to explain that the probability of inter-TAD contacts is only two-fold lower than intra-TAD contacts? Indeed, if only one enhancer-promoter contact is needed for gene activation, then a total insulation would not be possible. Alternatively, can one imagine a "phase-separation" model in which both the enhancer and the promoter would switch to a potentiated state as demixing proceeds? So far, no strong mechanistic conclusion can be drawn from the existing literature, and additional time-resolved experiments are needed.

3.3 Diffusion of transcription factors

3.3.1 Diffusion in a complex media

Most of the content is extracted from the review we wrote with Xavier: "Protein motion in the nucleus: from anomalous diffusion to weak interactions", published in Biochemical Society Transactions (Woringer and Darzacq 2018).

How exactly transient interactions are involved in the regulation of TF diffusion is unclear, but are reflected by live cell imaging techniques, including single-particle tracking (SPT). Overall, the macroscopic result of these microscopic interactions is almost always anomalous diffusion, a phenomenon widely studied and modeled.

Here, we review the connections between the anomalous diffusion of a TF observed by SPT and the microscopic organization of the nucleus, including recently described topologically associated domains and dynamic phase-separated compartments. We propose that anomalous diffusion found in

single particle tracking (SPT) data result from weak and transient interactions with dynamic nuclear substructures, and that SPT data analysis would benefit from a better description of such structures.

1. **Introduction.** Within this framework, the understanding of gene expression regulation reduces to elucidating how external factors (including TFs) affect the kinetic constants k. Although it can be assumed that kinetic rates are characterized only by the nature and concentration of enzyme, substrate and cofactors, it was shown in 1906 by Marian Smoluchowki (von Smoluchowski 1906) that the kinetic rate of a well-mixed, diffusion limited reaction can be decomposed as $k = 4\pi Da$. Thus, the kinetic rate k is a function of both the cross-section of interaction a (reflecting the chemical properties of the partners and usually studied by biochemical approaches) and the diffusion constant D of the species.

   Since D is determined by the local environment, this finding is striking in the context of gene expression regulation: now the kinetics of one reaction depend on the whole nuclear structure. More specifically, any factor that affects diffusion in any specific or non-specific way will ultimately influence reaction rates. Indeed, interactions resulting in facilitated diffusion on a substructure (such as a TF on DNA, Hammar, Leroy, et al. 2012; Normanno et al. 2015; Anders S. Hansen, Pustova, et al. 2017) or (transient) segregation inside a membrane-less compartment in a phase-separated manner (A. G. Larson et al. 2017; Strom et al. 2017) can all be seen under the unifying framework of diffusion on a surface of reduced dimensionality. Diffusion on surfaces of reduced dimensionality yields kinetics that are qualitatively different than in free, 3D diffusion and leads to potentially dramatically increased reaction rates (Kopelman 1988).

   Anomalous diffusion, a phenomenon occurring when a molecule explores a volume lower than predicted by diffusion, affects all proteins inside a cell. Numerous physical models can describe anomalous diffusion (Metzler, Jeon, et al. 2014), and several have been applied to the motion of nuclear proteins. However, many of them only provide a phenomenological description of diffusion, rather than mechanistic insights, and radically distinct models can often fit the available data equally well.

   In light of these considerations, it is worthwhile to examine recent discoveries describing either stable sub-nuclear compartments or their more transient, weak-interactions-induced counterparts to highlight their influence on the diffusion of factors through dimensionality reduction. This includes TADs (topologically associated domains), lamina-associated domains, nucleoli, non-coding RNAs, transcription factories, phase-separated domains, etc. They constitute substructures with a high valency amenable to weak interactions that can qualitatively influence diffusion and target search.

   Here, we first review anomalous diffusion models applied to protein motion and relate them to a potential physical generative model. Then, we emphasize recent advances in the characterization of regions of reduced dimensionality in mammalian nuclei, both aspecific through volume exclusion and specific through transient, weak-but-specific interactions. Finally, we propose that these weak interactions shape TF dynamics, and that single particle tracking (SPT) analysis would greatly benefit from a better understanding of the pairwise interaction map between nuclear proteins.

2. **Most anomalous diffusion models reflect underlying networks of weak interactions.** The technique of choice to investigate protein motion in the nucleus of live cells is light microscopy of fluorescently tagged proteins. Different imaging and modeling modalities have provided significant insights, including fluorescence recovery after photobleaching (FRAP), fluorescence correlation microscopy (FCS) or single-particle tracking (SPT). In this review, we focus on single-particle tracking, because it directly provides access to the dynamics of single diffusing molecules. In a SPT experiment, a small subset of the proteins of interest are imaged and tracked over short times at the resolution of a few tens of nanometers, allowing to resolve isolated single molecules even within clusters of high densities.

In solution, the diffusion coefficient D of a protein is inversely proportional to the hydrodynamic radius of the protein (r) and the viscosity of the medium (η) through the Stokes-Einstein relationship $D = \frac{k_B T}{6\pi\eta r}$ where $k_B T$ reflects thermal agitation, with k_B the Boltzmann constant and T the absolute temperature. This description, however, is too simplistic in the complex cellular environment. Indeed, with the exception of inert tracers of small molecular weight (Seksek, Biwersi, and A. S. Verkman 1997; M. Weiss et al. 2004), it is well acknowledged that (a) macromolecules in a cell diffuse much slower than in a medium of comparable viscosity, (b) that complexes of high molecular weight can diffuse faster than small proteins, and (c) that most molecules exhibit anomalous diffusion.

Thus, the diffusion of TFs cannot be described by simple friction/viscosity relationships, and their behavior, perhaps unsurprisingly, has to be seen from the angle of transient interactions with a dense matrix of interactants. In the context of this review, we define transient (or "weak") interactions as interactions that are usually too short-lived to be captured by traditional biochemistry techniques, that typically involve one or several wash step, during which proteins interacting specifically but transiently get diluted and washed out.

Furthermore, diffusion of many factors does not follow free, Brownian diffusion (Figure 3.4a,c). Such diffusion is termed anomalous (more specifically, sub-diffusive), meaning that the space explored over time by one factor is lower than expected by free diffusion (reviewed in Höfling and Franosch 2013; Metzler, Jeon, et al. 2014). Anomalous diffusion is usually characterized by a sub-linear growth of the mean squared displacement (MSD) as a function of time (that is, $MSD(t) \sim t^{\alpha}$ with $\alpha < 1$; Figure 3.4a) and α is called the anomalous diffusion exponent. This anomalous diffusion exponent does not fully characterizes the type of diffusion and additional metrics such as the distribution of translocation angles (Figure 3.4b) provide valuable information about the dynamics. Anomalous diffusion of proteins have been fitted with success by phenomenological anomalous diffusion models, and include fractional Brownian motion (Figure 3.4d) (Tejedor et al. 2010; Guigas and M. Weiss 2008; Shinkai et al. 2016; R. P. Ghosh et al. 2017), continuous time random walks (CTRW; Figure 3.4e) (Weissman, G. H. Weiss, and Havlin 1989; Saxton 2007a) and diffusion in fractal media (Figure 3.4f) (Ben-Avraham and Havlin 2000; Bancaud et al. 2009; Izeddin, Récamier, et al. 2014). Although useful as phenomenological descriptions, these models are often agnostic regarding the underlying reality of the process. In any case, the explanation of diffusion has to rely on physics and chemistry of the nucleus.

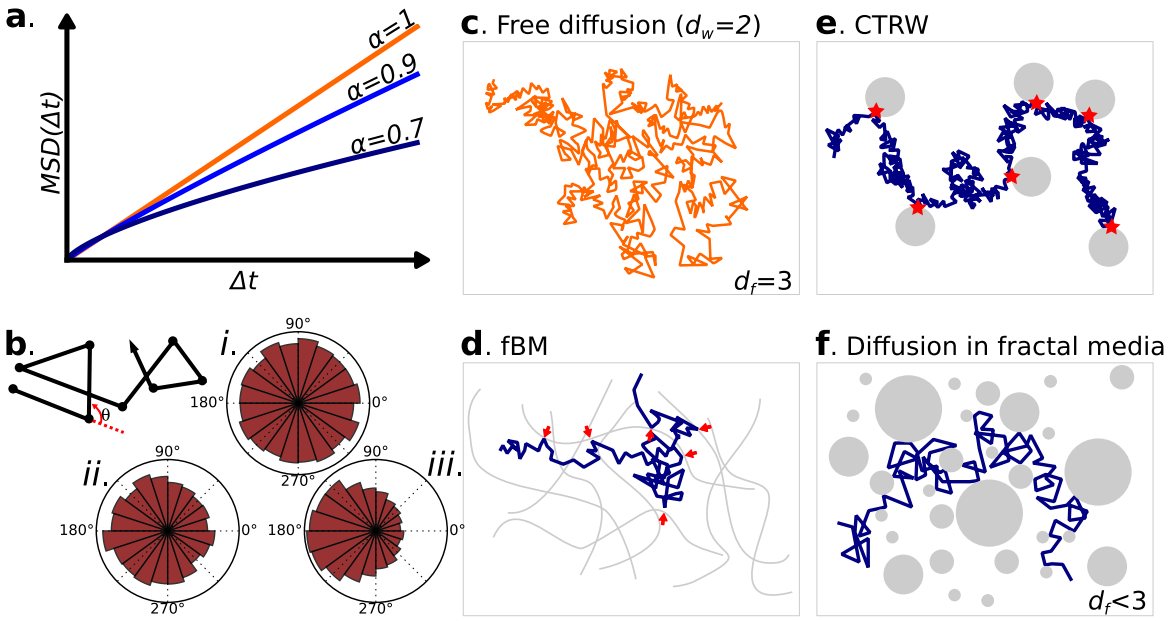


Figure 3.4: Models of anomalous diffusion and plausible underlying physical structures. (a) and (b) characterizations of anomalous diffusion. (a) Sub-linear mean-square displacement plotted as a function of time characterizes subdiffusion, and reflects how a diffusing particle explores space, the degree of anomalous diffusion is characterized by the exponent α, the lower the α the more sub-diffusive the process. (b) Free diffusion is characterized by isotropic distribution of angles – subpanel (i) – whereas anisotropic distribution indicates anomalous diffusion – subpanels (ii) and (iii) –. (c) 3D free diffusion ($d_w = 2$), as usually encountered in a homogeneous media ($d_f = 3$). (d)-(f) Several types of heterogeneous media can yield anomalous diffusion, including (d) diffusion within a viscoelastic polymer, in which a protein "bounces against" an elastic structure, a process traditionally described by fractional Brownian motion (fBm) and (e) free diffusion interspersed by long binding times –red stars–, a process called continuous time random walk –CTRW–, and (f) diffusion within a fractal media, that is a space obstructed by obstacles of all sizes.

From a physical perspective, proteins can adsorb and diffuse on nuclear substructures. When this happens, the exploration properties of the protein are universally given by two parameters: first, the dimension of the random walk d_w (deduced from the anomalous diffusion exponent: $d_w = \frac{2}{\alpha}$, i.e. the scaling of the MSD: $MSD(t) \sim t^{2/d_w}$), and second, the dimension of the space available to diffuse d_f. d_f can be integer ($d_f = 1$ for instance for sliding on DNA without jumps, $d_f = 2$ for a factor freely diffusing on the surface of a sub-compartment), or non-integer, a feature that characterizes self-similar structures, that is, fractals (Figure 3.4f). For instance, highly obstructed media or dense environments where the available volume is reduced to small pores are often accurately described by their non-integer fractal dimension d_f. Such a case can occur in dense sub-compartments such as phase-separated domains. For the sake of this review, we will denote structures of $d_f < 3$ as structures of *reduced dimensionality*.

Depending on d_f and d_w, the motion of the protein falls into two universal categories, termed compact and non-compact (de Gennes 1982; Condamin, Bénichou, et al. 2007; Bénichou et al. 2010). In a compact exploration ($d_w > d_f$), exploration is local and distance-dependent and a given site is explored repeatedly over time, in a highly recurrent manner. Conversely, in a non compact exploration ($d_w < d_f$), the exploration is global, and every site on the structure has a constant probability to be explored (distance independence); the exploration is non-recurrent (transient). For instance, a particle freely diffusing has a d_f of 2. When diffusion takes place in a 3D space ($d_f = 3$) the particle tends not to revisit sites, adopting a non-compact exploration (indeed, $d_w < d_f$). Conversely, a particle in free, Brownian diffusion ($d_w = 2$) constrained to

diffuse in 1D ($d_f = 1$; hypothetically along a DNA fiber) will repeatedly sample the same sites (compact exploration, $d_w > d_f$). Consequently, target search times are decreased and reaction rates are increased in the compact case.

Structures of reduced dimensionality, including fractals, emerge naturally from various processes, including diffusion-limited aggregation and hierarchical assembly of macro-molecular scaffolds, such as the multi-scale organization of chromatin. The goal of the next sections is to highlight a few structures of reduced dimensionality in the nucleus and how they influence kinetics of TFs.

3. **Steric hindrance in the nucleus.** Far from constituting a homogeneous medium, the nucleus is a highly organized and sub-compartmentalized organelle. The main organizing structure, chromatin, constitutes approximately 10-30% of the nuclear volume (Milo and Phillips 2016; Ou et al. 2017) and likely accounts for a significant part of the diffusion slowdown (Matsuda et al. 2014). Since every molecule has to slalom around a dense and heterogeneous chromatin environment, diffusion is impaired. Note that, however, similar diffusion coefficients are usually observed in the cytoplasm and the nucleoplasm (Guigas, Kalla, and M. Weiss 2007), suggesting that protein crowding can also account for diffusion slowdown (Ando and Skolnick 2010; McGuffee and Elcock 2010 and Bancaud et al. 2009 for a discussion).

   Over the past years, organizing principles of chromatin have emerged: at large scale, the genome is segregated in chromosome territories and regions of heterochromatin/euchromatin, lamina-associated domains at the periphery and nucleoli lying more at the center. At higher magnification, chromatin is organized in areas of preferential interactions such as A/B compartments and topologically associated domains (TADs) that reflect the functional organization of chromatin (Pueschel, Coraggio, and Meister 2016).

   Overall, although highly heterogeneous, chromatin in the mammalian nucleus is well described by a self-similar, fractal structure that occupies a non-zero volume. This was initially postulated (A. Grosberg et al. 1993), and later evidenced by spectroscopic (D. Lebedev et al. 2005; D. V. Lebedev et al. 2008), genomic (Lieberman-Aiden et al. 2009; A. Yu. Grosberg 2012) and imaging techniques (Bancaud et al. 2009; Récamier et al. 2014; Shinkai et al. 2016; Shinkai et al. 2017; S. Wang et al. 2016). Note that this description as a fractal, however, cannot be exact, even from a mathematical standpoint, and some properties usually applicable to fractals might not be relevant for chromatin (K. Huang, Backman, and Szleifer 2018).

   As a consequence, factors diffusing in the available volume are constrained by this structure (Goulian and Simon 2000), possibly experiencing diffusion in a medium of reduced dimensionality, as evidenced by numerous reports (Bancaud et al. 2009; Anders S. Hansen, Woringer, et al. 2018). In a model where only volume exclusion happens, proteins of the same size and shape should have the same diffusion coefficient. Thus, the embedding structure of the nucleus only sets a lower bound on the level of anomalous diffusion that can be observed.

   Several lines of argument point to the fact that although steric hindrance induces a significant decrease in the diffusion coefficient (D), its effect on anomalous diffusion (α) is mild for proteins of weight < 150 kDa (that is, about the weight of a histone octamer wrapped with 150 bp DNA). Indeed, FRAP experiments performed with protein or non-protein tracers of increasing molecular weights suggest that low molecular weight tracers diffuse almost freely in the nucleus, allowing to infer a viscosity close to the one of water (Seksek, Biwersi, and A. S. Verkman 1997). At molecular weights > 150 kDa, anomalous diffusion becomes more prominent (Guigas, Kalla, and M. Weiss 2007; M. Weiss et al. 2004; Bancaud et al. 2009), sometimes leading to particles occasionally being trapped in the chromatin mesh. This effect is consistent with the relatively limited volume occupied by chromatin (Milo and Phillips 2016; Ou et al. 2017), that can trap big complexes but not smaller proteins.

   In conclusion, although volume exclusion by chromatin and other nuclear constituents is real, it affects all proteins of the same size in a similar manner. In contrast, a protein weakly but

specifically interacting with such a structure (for instance, TFs sliding/hopping on DNA; Wunderlich and Leonid A. Mirny 2008; Hammar, Leroy, et al. 2012; Coppey et al. 2004; Iwahara, Zweckstetter, and Clore 2006; Doucleff and Clore 2008) will immediately show a much higher level of anomalous diffusion. Furthermore, even without considering a fractal structure, simple dimensionality reduction to 1D or 2D can yield non-traditional kinetics termed fractal kinetics (Kopelman 1986; Kopelman 1988; Berry 2002). Fractal kinetics are characterized by a time-dependent reaction rate: for a $A+A \rightarrow$ products reaction, the reaction rate k_{app} can be expressed as $k_{\text{app}} = kt^{-h}$, with h that depends on the (fractal) dimension of the space. Fractal kinetics arise when the reaction cannot be assumed to be well-stirred, which is the case of diffusion on surfaces of reduced dimensionality. For instance, fractal kinetics in 2D could occur by weak interaction with the nuclear lamina, Figure 3.4b. All in all, weak and transient interactions shape the nuclear landscape and can give rise to emergent structures and properties, as exemplified in the next section.

Even though live imaging approaches specifically characterize the behavior of one single factor, they are blind to all these substructures. Indeed, SPT reflects the dynamics of proteins transiently interacting with those structures of reduced dimensionality and one TF potentially visits several of them in the span of a few tens of milliseconds (S. K. Ghosh, Cherstvy, and Metzler 2015). Such complex behavior therefore appears macroscopically as various kinds of anomalous diffusion.

4. **Perspectives: seeing beyond the dots.** In a complex mammalian nucleus, the diffusion of a TF is ruled by transient interactions with underlying structures of reduced dimensionality. From a more general perspective, the question arises of how gene expression regulation processes relate to the multiplicity of structures of reduced dimensionality?

   Proteins often harbor several domains (Figure 3.5), holding the potential to interact successively and repeatedly with multiple classes of structures of reduced dimensionality. Thus, depending on its interaction domains, a TF will "see" a different landscape and will interact with some structures whereas other factors will either be excluded or cross them without any additional interactions than limited steric hindrance. In this respect, the nucleus can be described as a "multiverse", in which some factors coexist in the same physical space but exhibit radically distinct dynamics and interactions (Figure 3.5).

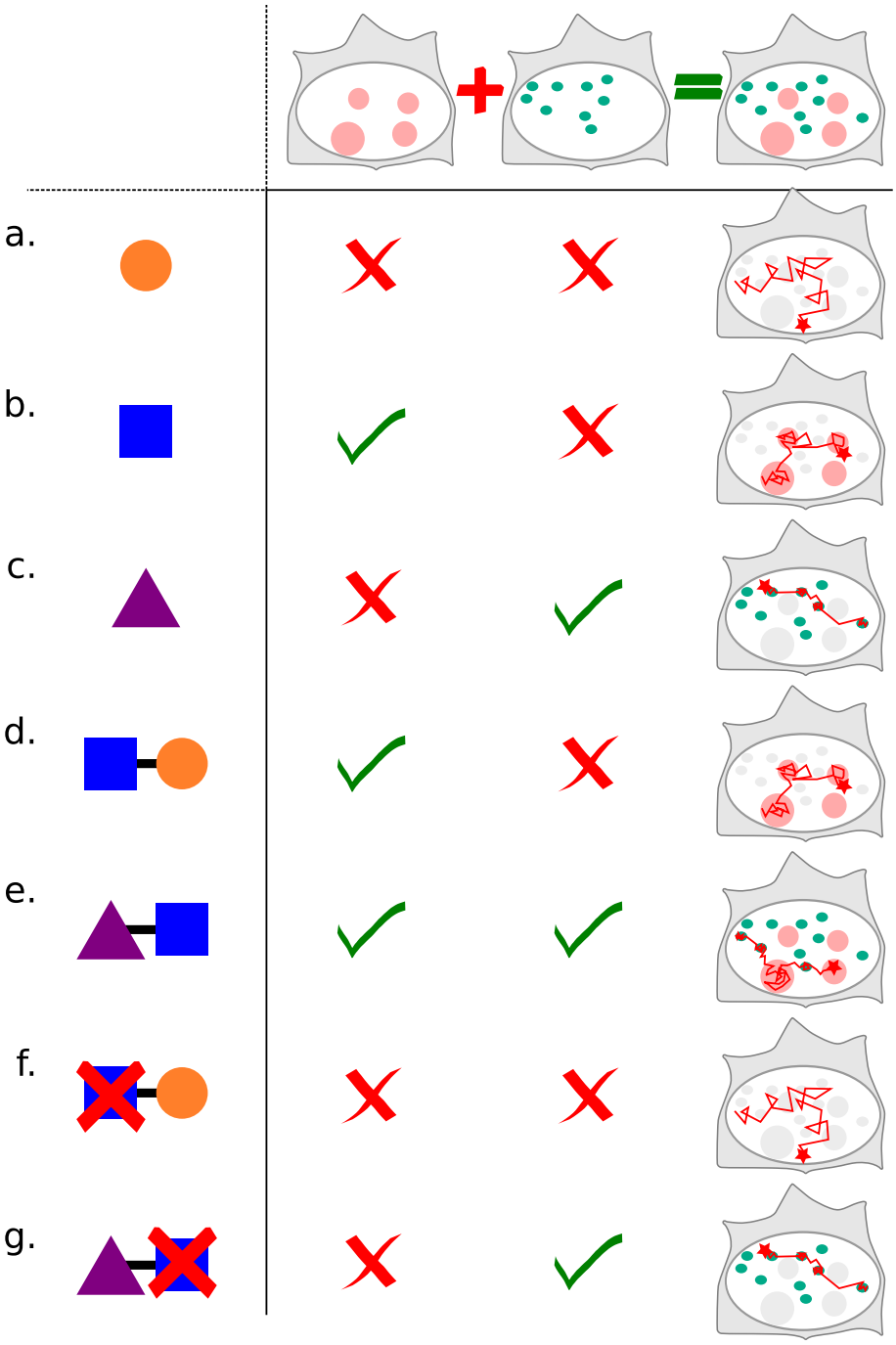


Figure 3.5: Weak interactions of individual protein domains shape TF dynamics. (a)-(c) individual protein domains have specific transient interactions: the round domain (a) does not interact with any particular structure (represented by the two columns of the table), (b) the square domain interacts with a given pink structure (first column), and (c) the triangle domain interacts with the green structures (second column). This results in domain-specific dynamics (third column). (d-e) When protein domains are combined within a protein or TF, the observed SPT is a mixture between the interactions of each single domain. (f-g) When individual domains are mutated, the protein loses some transient specific interactions, and its dynamics can dramatically change (compare panels e and g).

Furthermore, structures of reduced dimensionality have been proven to be functionally relevant. For example, the dynamic and regulated switching of a TF between structures of reduced dimensionality determines its dynamics (Figure 3.5a-e) and function. It has been shown that TF

exhibit radically different dynamics before/after a post-translational modification (Loffreda et al. 2017; Speil et al. 2011), or an artificial deletion of a domain (Figure 3.5f-g), (Anders S. Hansen, Pustova, et al. 2017; J. Elf, G.-W. Li, and X. S. Xie 2007; D. Mazza et al. 2012; Sekiya et al. 2009; Clauß et al. 2017; Zhen et al. 2016). In that case, the observed diffusion will be arising from the remaining interactions from the other interaction domains, or ultimately from simple volume exclusion (Isaacson, McQueen, and Peskin 2011).

Although theoretical and experimental support for the importance of weak interactions as an architectural principle of the nucleus and gene expression regulation is being actively investigated, several questions remain unaddressed:

First, how many distinct types of structures of reduced dimensionality exist? Since the numbers of types of low-complexity domains is likely to be limited, one can expect that a limited number of such structures actually coexist at a given time in a nucleus (Rahul K Das, Ruff, and Rohit V Pappu 2015). This implies that the SPT dynamics of TFs will fall in a limited number of categories, which in turn is determined by their combinatorial interactions with one or several of these structures. To take into consideration such processes paves the way for a higher-order understanding of gene expression regulation and key transitions occurring for instance during mitosis or development.

Second, can we determine the pairwise interaction matrix between low-complexity protein domains, which would allow to derive predictive dynamics of a given TF modification? Ideally, such matrix will encompass all known low-complexity domains, but also abundant multivalent RNAs and DNA sequences, and each element of this matrix will reflect the affinity between two domains under physiological conditions (Figure 2.1d).

Third, how much detail is required to describe these structures of reduced dimensionality? Is the pairwise interaction between protein domains a good approximation of the properties of the nucleus? Conversely, one can imagine substructures of reduced dimensionality arising from interactions more complex than simple pairwise-interactions. Indeed, it is widely known that cooperativity plays a role in the assembly of many more or less stable macromolecular structures (Chronis et al. 2017), including some phase-separated domains (Pak et al. 2016).

Fourth, how do key biological transitions such as differentiation intertwine with these structures of reduced dimensionality? In a similar way as pluripotency or cell-type specific TF networks have been identified, can pluripotency or cell-type specific structures of reduced dimensionality be evidenced, integrating the expression levels of TFs and providing a framework to better understand such key processes?

To answer those questions, our understanding of nuclear processes need to be drastically expanded. Hitherto, a dynamic picture of spatially segregated factors, together with their interaction matrix, is currently missing. Promising tools to access those parameters include quantitative FRET (Sukenik, Ren, and Gruebele 2017), in cell NMR (Maldonado, Burz, and Shekhtman 2011; Freedberg and Selenko 2014; Theillet et al. 2016), low-photons SPT (Balzarotti et al. 2017), tracking FCS (Limouse et al. 2017), spatially resolved FCS (A. P. Singh et al. 2017) and computational methods (Quiroz and Chilkoti 2015; Harmon et al. 2016).

Conclusion.

Although the so far identified key players in gene expression regulation are biochemically stable complexes that can be purified using traditional methods, increasing evidence suggest that higher-order, weaker-interaction structures, acting as structures of reduced dimensionality, play a central role in transcriptional regulation. They do so by providing a remarkably versatile way of specifically and timely regulating TF target search dynamics and thus gene expression. All in all, it appears that the functional properties of the nucleus are not only shaped by stable macromolecular complexes but also transient structures (arising from a continuum of weak interactions

that might seem spurious). In this context, the saying from Heraclitus makes probably more sense than ever: "The fairest order in the world is alike a heap of random sweepings".

Glossary.

Anomalous diffusion: a phenomenon occurring when a molecule explores over time a volume lower than predicted by Brownian diffusion. It is usually characterized by a sub-linear growth of the mean squared displacement (MSD) as a function of time ($MSD(t) \sim t^{\alpha}$ with $\alpha < 1$), with α the anomalous diffusion exponent.

Fee diffusion: (also termed Brownian diffusion) characterizes the motion of a particle in a fluid arising from thermal agitation only. Seen as a "null" model in SPT.

Phase separation: a state of matter in which part of the soluble protein fraction segregates into a liquid or liquid-like droplet.

Fractals: structure exhibiting (statistical) self-similarity, that the (statistical properties of the) structure remain similar at various zoom levels. Fractals can be described by their (potentially non-integer) fractal dimension d_f. The fractal dimension of an object describes its space-filling property. Many types of biological objects exhibit fractal properties, such as the branching pattern of the lung, or the hierarchical folding of DNA in the nucleus.

Weak interactions: (in this review) interactions that are usually too short-lived to be captured by traditional biochemistry techniques, that typically involve one or several wash step, during which proteins interacting specifically but transiently get diluted and washed out.

Facilitated diffusion: biophysical phenomena allowing a molecule to find its target faster than predicted by traditional, 3D free diffusion. This includes diffusion on a surface of reduced dimensionality such as DNA.

Surface of reduced dimensionality: an object whose fractal dimension $d_f < 3$, meaning that it exhibits some properties similar from the ones of 1D or 2D structures.

Compact exploration: universal mode of diffusion in which the exploration of the diffusing molecule is local and distance-dependent and a given site is explored repeatedly over time, in a highly recurrent manner. Within the compact exploration mode, several strengths of compaction can be distinguished.

Non-compact exploration: universal mode of diffusion in which the exploration is global, and every site on the structure has a constant probability to be explored (distance independence).

Fractal kinetics: type of kinetic reactions happening within a reactor that is not well-stirred. This notably includes most reactions happening on a surface of reduced dimensionality. Fractal kinetics are characterized by a progressive segregation between reagents and products. The kinetic rate k of the reaction is time-dependent

3.3.2 Overview of the diffusion of known TFs

In the previous sections, we progressively detailed how the nuclear environment is organized (section I.2), and how transcription proceeds and is regulated (section 3.1). We then motivated the importance of studying how TFs diffuse in the cell to understand how transcription is regulated (section 3.3.1). We now turn to a review of the current knowledge of TF dynamics. Indeed, the nuclear dynamics of several TFs has been studied, and we highlight here some of the findings that were obtained.

1. **Most TFs diffuse relatively slowly compared to the viscosity of the medium.** A first series of results looked at the diffusing fraction of TFs in the nucleus. Compared to their expected diffusion coefficient based on the viscosity of the medium (section 3.3.1), most TFs exhibit a reduced diffusion coefficient by at least an order of magnitude.

(a) *Common Mode of Transcription Factor Interaction with Chromatin.* For instance, the pluripotency factor Sox2 has been shown to exhibit differential diffusion coefficients depending on which sub-nuclear compartment is observed. In the whole nucleus, 62% of the Sox2 protein diffuses at a diffusion coefficient of 2.7 $\mu m^2/s$ (the rest being bound to DNA), whereas in enhancer clusters, 36% of the proteins only are diffusing and diffuse at an even reduced diffusion coefficient (1.4 $\mu m^2/s$). In heterochromatin (HP1 domains), Sox2 exhibits 3 fractions, one bound fraction (16%), a slow fraction (D=0.6$\mu m^2/s$; 26%) and a fast fraction (1.6 $\mu m^2/s$; 58%; Z. Liu, Legant, et al. 2014).

On a more general note, most of the TFs ever observed display a diffusion coefficient around 1-10 $\mu m^2/s$. For instance the glucocorticoid receptor (GR) have been shown to diffuse at 3.4-5.2 $\mu m^2/s$ (Mueller, Wach, and McNally 2008; J. Christof M. Gebhardt et al. 2013), p53 diffuses around 4.3-4.4 $\mu m^2/s$, Max around 8.0 $\mu m^2/s$ (Mueller, Wach, and McNally 2008), SRF (a PRC2 component) displays three noticeable fractions (35% bound, a slow fraction diffusing at 2.18 $\mu m^2/s$ and fast fraction diffusing at 10.82 $\mu m^2/s$; Hipp et al. 2019). PRC2 also displays two fractions (20% bound; 80% free diffusing at 2.09 $\mu m^2/s$; Youmans, Schmidt, and Cech 2018). When inserted in mammalian cells, the bacterial TF TetR also displays three sub-populations (24% slow, $<$0.1 $\mu m^2/s$; 43% medium at 1 $\mu m^2/s$ and 33% diffusing fast, around 8 $\mu m^2/s$; Normanno et al. 2015).

All these findings confirmed a foundatory analysis (Mueller, Wach, and McNally 2008) that highlighted the similarities in the dynamics of several transcription factors, suggesting common modes of exploration of the nucleus and of DNA-binding.

(b) *A database of dynamic parameters of transcription factors.* To undertake a more systematic analysis, Stephan Ortiz, under the supervision of Mustafa Mir, and with the help of Anders Hansen, started to compile an exhaustive list of published TF dynamics measurements, with the goal to provide an integrated database about the dynamics of TFs, but also to easily perform some meta-analyses. I provided some help to compile this database, that now contains several hundreds of entries.

2. **Most TFs bind DNA for a few seconds.** A second "general rule" learned from single-molecule imaging is the fact that despite being crucial for gene activation, most TFs bind DNA for only a few seconds.

(a) *Most TFs contact DNA for short periods of time...* This has been evidenced for many factors, including Sox2, that has a residence time of $\sim$ 7s, with $\sim$ 88% with a residence time $<$ 3s (Z. Liu, Legant, et al. 2014). This estimate was then split in two sub-populations of bound molecules, with a "fast binding" faction (binding in average less than a second; 0.8 s) and a "long binding" fraction (12s in average; J. Chen et al. 2014). p53 bound fraction can also be divided between a long binding (6.5 s; 75%) and a fast-binding (1.8 s; 25%; Loffreda et al. 2017). Similarly, NFκB displays a short binding fraction (0.53s; 96%) and a long binding fraction (4.1s; 4%; Bosisio et al. 2006; Callegari et al. 2018). Moreover, p53 and GR display distinct residence time when they are closed to active genes (Morisaki et al. 2014) Even pioneer factors such as FoxA1 display very similar dynamics as non-pioneer factors (Swinstead et al. 2016).

(b) *But there are some exception.* Among the few exceptions are SRF, a TF whose residence time is around 1 min (Hipp et al. 2019). This residence time is exceptionally long for a TF, but common for other DNA binding proteins such as histones (that reside on DNA for hours) or CTCF (Anders S. Hansen, Pustova, et al. 2017). Also, the TFIID core protein TBP binds DNA for seconds to minutes, an estimate that was provided by several techniques, including SPT in various models (Reisser and J. Christof M. Gebhardt 2017; Teves, An, Bhargava-Shah, et al. 2018) and temporal ChIP (Poorey et al. 2013).

(c) *Facilitated dissociation.* In a simple kinetic model, the time spent on chromatin follows a simple exponential law: it assumes a first order kinetic. If two types of sites (resp. long and short residence times) are observed simultaneously, the residence time estimate will be a mixture of several exponentials. Under this model, the residence time only depends on the affinity of a given TF for the specific DNA sequence.

However, under some conditions, dissociation from DNA can be related to a non-first order reaction: a phenomenon termed facilitated dissociation. Under this model, freely floating proteins transiently associate with a bound molecule and destabilize it. Under this model, the residence time of the protein depends on the concentration of external factor, which is not the case under first-order kinetics. Facilitated dissociation has been initially observed in bulk for the bacterial TF Fis (Giuntoli et al. 2015), then simulated using molecular dynamics (M.-Y. Tsai et al. 2016) and finally observed at the single-molecule level (Kamar et al. 2017).

(d) *Hopping.* Finally, when a TF dissociates, it might either diffuse freely in 3D for a long time, or quickly re-associate with chromatin, a phenomenon termed hopping, that is particularly enhanced in high density chromatin, as evidenced by a series of theoretical models (Wunderlich and Leonid A. Mirny 2008; Cortini and Filion 2018; Amitai 2018). Some of them predicted increased gene expression at some loci (Avcu and Molina 2016). From an experimental perspective, several NMR studies have shown that TF hopping and segment inter-transfer are present, at least in an *in vitro* setting (Iwahara, Zweckstetter, and Clore 2006; Doucleff and Clore 2008; Speil et al. 2011).

3. **TF dynamics are regulated.** Second, emerging experiments show that the dynamics of a TF is not an intrinsic property, but that it can rather be regulated. For instance, the acetylation of the C-terminal domain of p53 correlates with its residence time (Loffreda et al. 2017).

   Another series of experiments comes from the observation of the dynamics of TBP. In Zebrafish, the bound fraction of TBP increases across development, a phenomenon that correlates with the reduced cell volume (Reisser and J. Christof M. Gebhardt 2017). During mESC mitosis, the bound fraction of TBP decreases from 34.6% in interphase to 18.6%. The diffusion coefficient of TBP also drops from 3.2 $\mu m^2/s$ in interphase to 2.4 $\mu m^2/s$ in mitosis. TBP experiments show how there can be an interplay between chromosome structure and TF dynamics.

   Finally, not only the diffusion coefficient and bound fraction of a protein can be modified, but also the way it explores space (section 3.3.1). Indeed, (Izeddin, Récamier, et al. 2014 and Bosanac et al, unpublished data) found that the P-TEFb transcription factor switches from a compact to a non-compact mode of exploration when interaction with the 7SK RNA is impaired. These findings can be explained theoretically, and a model linking the mode of exploration with bursting parameters has been proposed (Meyer et al. 2012).

4. **Detecting co-recruitment and oligomerization states.** Single-molecule dynamics can also be used to assess protein-protein interactions. We just highlight two examples. First, (Anders S. Hansen, Pustova, et al. 2017) measured distinct diffusion coefficients for CTCF and cohesin, which lead to the conclusion that the two complexes could not be diffusing together in the nucleus.

   Another example looked at HP1 oligomerization dynamics using fluorescence correlation spectroscopy (FCS), and found that HP1 diffuses as a dimer and tetramerizes upon DNA binding (Hinde, Cardarelli, and Gratton 2015).

5. **The long quest for specific DNA binding.** The residence time of most TFs can be decomposed in two main components, a "fast" and a "slow" component, which lead the authors to propose that the "slow" (long residence time) sub-population might represent specific binding of a TF with a cognate TF binding site while the short might be transient, non-specific interactions.

This vision is usually supported by experiments deleting the DNA-binding domain and observing changes in the "slow" fraction. However, this relationship between the duration of binding and specificity has not been entirely demonstrated, and a clear way of determining a specific vs. nonspecific contact is not available yet, as initially discussed in (Mueller, Stasevich, et al. 2013).

In some contexts (TF used in an exogenous context where no specific binding site exists), non-specific binding has been measured and shown to be of the order of 100-200 ms (159 ms for TetR transfected in mammalian cells; Normanno et al. 2015, 182 ms for LacI in mammalian cells; Caccianini et al. 2015), supporting the idea that non-specific interactions are usually short-lived.

However, some recent experiments challenged this view recently. First, (McSwiggen et al. 2018), measured the dynamics of RNA PolII when it is hijacked to transcribe the HIV rather than the host genome. They found that despite the fact that the HIV is not particularly enriched in PolII motifs, it gets almost entirely recruited to the HIV genome. The authors suggest that this happens because the HIV genome is artificially maintained nucleosome-free. Within the framework of specific/unspecific interactions, PolII is having mostly non-specific interactions with the HIV genome, while spending a lot of time on chromatin, which contradicts the initial guess that non-specific interactions are short-lived.

Second, (Raccaud et al. 2018) measured the link between mitotic binding of TFs a properties and their properties in interphase dynamics. They found very little correlation between the number of ChIP-seq peaks (supposedly related to specific interactions) and the dynamics of the TF (with the exception of the pseudo on-rate), further suggesting that disentangling specific from non-specific binding using only dynamics remains an open challenge.

Chapter 4

Imaging approaches to study nuclear diffusion

Introduction In the previous sections, we highlighted the interplay between nuclear organization, TF dynamics and transcription regulation and demonstrated the need to perform live-cell imaging in order to better understand transcription regulation.

In this section, we review existing imaging and analysis techniques that were used to image transcription-related processes. We put a special emphasis on single-particle tracking (SPT). In section 4.1, we detail the traditional workflow of a SPT experiment. In section 4.1.3, we provide an overview of other complementary imaging techniques. We then review in section 4.2 some mathematical basics behind the analysis of diffusion and in section 4.4, we provide a critical comparison of SPT analysis techniques.

4.1 Single-Particle tracking

Single-particle tracking (SPT) is a fluorescence, live cell, single-molecule technique that follows across time individual molecules (see Z. Liu, L. Lavis, and Eric Betzig 2015 for a comprehensive review). In practice (detailed in section 4.1.1), proteins of interest are labeled and imaged at low density using a high sensitivity microscope. Computational analysis then allows to extract detections and to reconstruct the track of each single molecule observed. Once the tracks are recovered, several analysis techniques can be implemented (section 4.4)

4.1.1 A standard pipeline to perform SPT

In this section, we present a traditional SPT experiment. We briefly review some key steps of SPT, from the preparation of the cell to data analysis.

1. **Fluorophores & labeling.** First, the protein of interest has to be labeled. Several parameters should be taken into account at that step:

   (a) *Choice of the labeling technique.* In order to perform live cell imaging, the protein is most of the time fused to either a fluorescent protein or a reactive tag (such as a HaloTag or a SNAP-tag So, Yao, and J. Rao 2008). Alternative approaches include imaging with nanobodies (small antibody-like peptides) or ad-hoc reporters (Tanenbaum et al. 2014; Delachat et al. 2018).

   When opting for a fusion protein, one can either transfect a construct, generate a genome-edited cell line that would carry the fusion protein at the endogenous locus, or use micro-injection to directly inject labeled proteins. Transient transfections are simple to perform but give poor control over the expression level. Genomic knock-ins are tedious to perform,

even with CRISPR-Cas9, but can fully reproduce the expression level and regulation of the endogenous protein. Micro-injection allows to use direct labeling of the protein with the dye, without the synthesis of a heavy fusion protein.

To obtain a sufficiently low number of visible fluorescent proteins at a time, one can rely on either a transfection with a low amount of plasmid, the labeling with a low concentration of dye, or stochastic labeling (Hui Liu et al. 2017)

(b) *Choice of the fluorescent protein or dye.* When using a fluorescent protein, brightness and photo-stability are usually the limiting factors to perform SPT. Furthermore, the choice of a photo-convertible or photo-activatable protein greatly facilitates imaging (van de Linde and Sauer 2014). Such proteins include Dendra2 (Gurskaya et al. 2006) and mEOS (Paez-Segala et al. 2015). When using a conjugation with a reactive tag, the choice of the synthetic dye is also crucial. Limiting factors are usually the cell permeability of the dye, and their cytotoxicity (B. R. Martin et al. 2005; Yang et al. 2011; D. S. Liu et al. 2014; Butkevich et al. 2017; R. P. Ghosh et al. 2017; Jonathan B. Grimm et al. 2017; Kolmakov et al. 2015).

To choose a fluorophore is usually complicated, and few objective comparisons of fluorescent proteins are available in the literature. However, some recent preprints provided a few hints. Dunsing et al. 2018 benchmarked fluorophore pairs to assess the fraction of proteins that actually emit fluorescence. Balleza, J. M. Kim, and Cluzel 2017 evaluated the maturation time of several fluorescent proteins. Banaz, Makela, and Uphoff 2018 compared fluorescence proteins and organic dyes with a focus on SPT.

2. **Imaging.** Imaging is then performed on live cells using a high sensitivity microscope equipped with high-power laser. Several critical parameters should be considered when performing imaging:

   (a) *Exposure time.* The faster the imaged protein moves, the faster one should image, and SPT of nuclear factor is usually performed at 5–10 ms exposure time. Furthermore, in order to avoid motion blur (detailed in section 1.2), one might want to use exposure strobes that are shorter than the acquisition time of the camera, a technique called stroboscopic illumination (reviewed in section 4.1.2).

   (b) *Signal-to-noise ratio.* Second, SPT is a single molecule technique, and the signal of a specific protein can quickly be overwhelmed by out-of-focus background fluorescence. A good signal to noise ratio (SNR) is ensured by the use of low-auto-fluorescence samples with clean coverslip and a proper choice of wavelength (cells are strongly auto-fluorescent in the blue-green region of the spectrum). Furthermore, imaging techniques that provide higher axial sectioning improve the SNR. Such techniques include TIRF, HiLO and light sheet techniques (Tokunaga, Imamoto, and Sakata-Sogawa 2008,Greiss et al. 2016; Tang and K. Y. Han 2018).

3. **Detection & tracking.** Once imaging has been performed and images of single molecules collected, the next step is to extract the localization of the molecules and to link them to form trajectories.

   (a) *Detection.* The literature on spot detection and single-molecule detection is extensive, and many of the advances from the PALM/STORM community can be readily used for SPT. Detection algorithms include ThunderSTORM (Ovesnỳ 2016), QuickPALM (Henriques et al. 2010), MTT (Sergé et al. 2008) and Picasso (Schnitzbauer, Strauss, et al. 2017). Many others are presented in (R. Sun, Archer, and Paninski 2016; Gustafsson et al. 2016; Bernhem and Brismar 2017; Przybylski et al. 2017; Boyd et al. 2018) and reviewed in (Sage, Pham, et al. 2018).

   (b) *Tracking.* Tracking, especially in high density conditions, has been the subject of intensive research, including a tracking challenge to benchmark existing algorithms (Chenouard et

al. 2014). The longer the track, the more critical the tracking algorithm (Saxton 2008). Examples of tracking algorithms include (Karrenbauer and Wöll 2013; I. Sbalzarini and Koumoutsakos 2005; Jaqaman et al. 2008; Sergé et al. 2008; Matysik and Kraut 2014; R. Sun and Paninski 2018; El Beheiry, Türkcan, et al. 2016). Most of them rely on linear programming and discrete optimization in order to split/merge tracks when ambiguities arise due to spots being close from one another. It is important to note that despite intense efforts to reconstruct tracks in a high density setting, some level of ambiguity will always remain, potentially hampering dramatically downstream analyses. As such, low density imaging conditions are preferable for SPT (Anders S. Hansen, Woringer, et al. 2018).

4. **Formats.** Once tracks have been extracted from the images, such results can be stored in a file format suitable for analysis and sharing. The literature is relatively poor, and little standards exist. Nonetheless, a SPT format based on the JSON format was proposed in Greenfeld et al. 2015. This format has not been maintained in the past four years, and this is why we decided to propose a new SPT standard (section 4.4).

5. **Conventional metrics and analysis.** Depending on the analysis technique, SPT traditionally yields some information on the existence of sub-populations of diffusing molecules and their respective diffusion coefficients (D). It also provides information on the residence time on DNA spent. To extract these parameters and many others, various analysis algorithms have been proposed. We review some of them in section 4.4.

4.1.2 Techniques to perform SPT

We now detail a few flavours of imaging techniques, that were designed to provide different types of data. We highlight some of their assets and drawbacks. Many questions can be answered using SPT (Figure 4.1), and include transcription-related questions and many others.

1. **spaSPT and other "fast SPT".** A first approach relies on fast/short exposure time such that individual molecules can be accurately followed over time. It is hard to tell when the technique was pioneered, but SPT inside cells seems to date back to 2000-2008 (Goulian and Simon 2000; Grünwald, Hoekstra, et al. 2006; Grünwald, R. M. Martin, et al. 2008). "fast SPT" is the "traditional" SPT, used in many studies to study transcription factors dynamics (see for instance Davide Mazza, Ganguly, and McNally 2013; Izeddin, Récamier, et al. 2014; Miller et al. 2017; Anders S. Hansen, Pustova, et al. 2017; Anders S. Hansen, Woringer, et al. 2018; Anders S Hansen et al. 2018). "fast SPT" has also been used to study non-transcription related processes, such as DNA repair in bacteria and on DNA curtains (Gorman et al. 2012; Stracy et al. 2016), the dynamics of the Cas9 protein in mammalian cells (Knight et al. 2015), the spatially-resolved switching dynamics of RNA-binding proteins in *C. elegans* (Wu et al. 2018), the dynamics of telomerase subunits (Schmidt, Zaug, and Cech 2016) or chromatin itself (Shinkai et al. 2017; Nozaki et al. 2017).

   Significant effort has been devoted in order to validate that SPT provides similar estimates as other techniques, such as FRAP and FCS (D. Mazza et al. 2012; Davide Mazza, Mueller, et al. 2013).

   Using this technique, the proteins are illuminated to maximize the quality of the tracking of fast moving particles. This implies the use of (a) fast exposure, typically 5-10 ms, and potentially stroboscopic illumination (b) high laser power, to be able to detect molecules despite the short exposure time. Because high laser power is used and because fast molecules go out-of-focus quickly, the obtained trajectories are usually short, which hampers correlation analyses and residence time analyses.

2. **Slow SPT.** "slow SPT" was designed to focus on DNA-bound molecules. By adopting a long exposure time, fast moving (non-DNA bound) molecules undergo motion blur and only immobile

proteins appear as detectable spots. A typical exposure time is 500 ms, although more complex exposure schemes have been proposed (Reisser and J. Christof M. Gebhardt 2017; Hettich and J Christof M Gebhardt 2018). The technique was pioneered in (J. Elf, G.-W. Li, and X. S. Xie 2007), and further used to characterize the residence time of many factors on DNA (see for instance Etheridge et al. 2014; J. Chen et al. 2014). This technique was also used to suggest that facilitated diffusion on DNA occurs in cells (Hammar, Leroy, et al. 2012) and to dissect complex kinetics of transcriptional regulation (Hammar, Walldén, et al. 2014).

3. **3D SPT.** Performing 3D tracking is comparatively harder than performing 2D SPT. We highlight some of the reasons in section 1.2. Nonetheless, at least one study performed 3D SPT in cells using PSF shaping (Thompson et al. 2010). This study focused on mRNA tracking. We build on this idea to propose a protein 3D SPT scheme in section 1.2.

4. **Other techniques.** Finally, a few other techniques disrupted the traditional trade-off between quality of the SPT data and quantity of data that could be acquired. A first one (Tsunoyama et al. 2018) performed regular fast SPT using an oxygen-scavenging system to increase the stability of the fluorophore.

   A second one, (Balzarotti et al. 2017) used a mechanical closed-loop feedback system to track one fluorophore at a time using a donut-shaped PSF: the diffusing molecule is properly tracked when the protein is inside the donut (in the dark space). Thus, the fluorophore is not excited and has a very low probability of photobleaching. When the protein moves, it enters the donut of light and emits photons. The collected photons are used to adjust the feedback system and move the donut so that the protein is again inside the donut. In this system, the motion of the donut reflects the location of the protein.

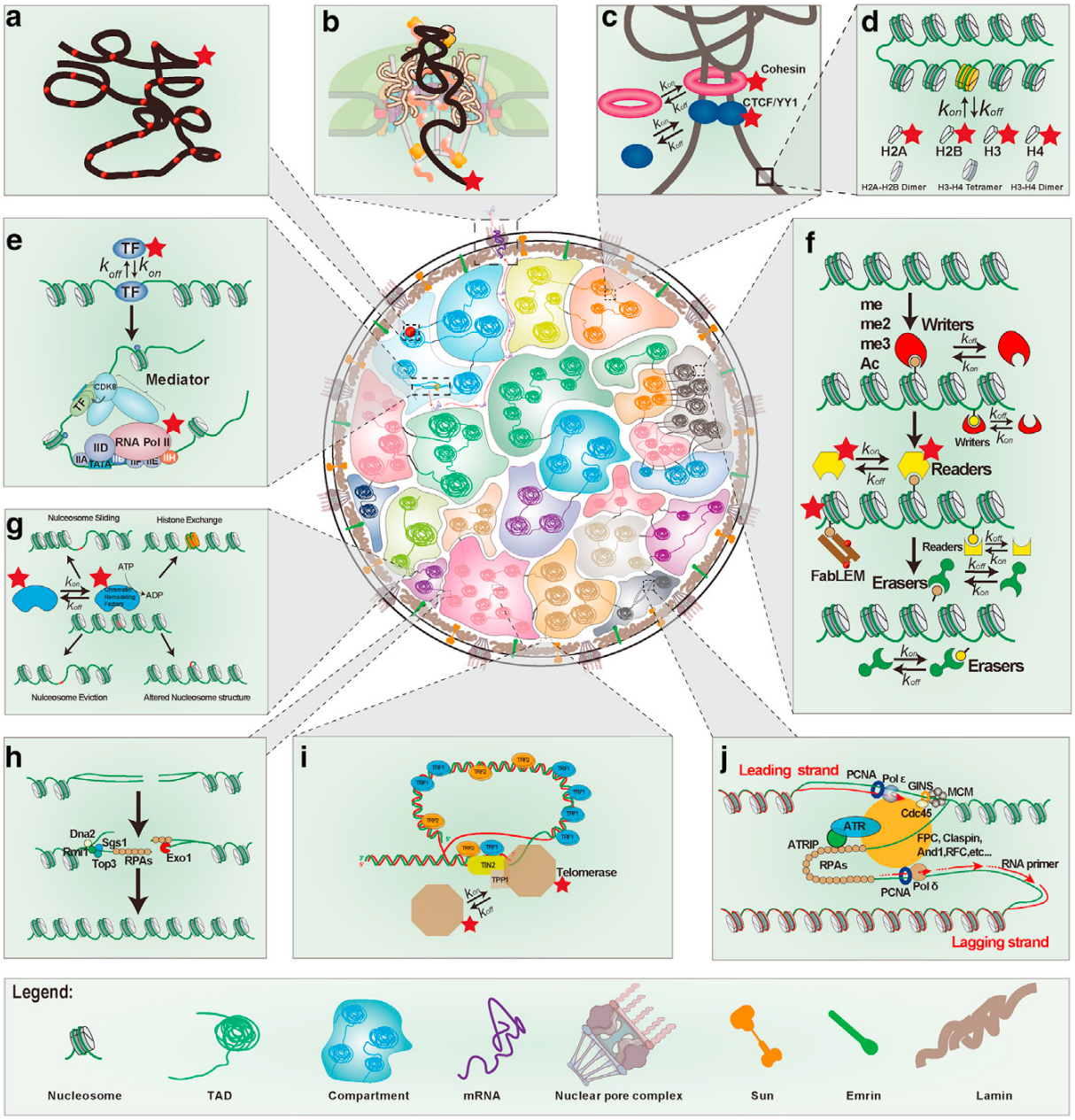


Figure 4.1: **Biological objects in the nucleus that can benefit from single-molecule imaging for mechanistic studies**. The red-star-labeled objects indicate objects that have been studied by single-molecule imaging. **(a)** The labeling of the chromatin locus is shown. Chromatin loci in different states (cell cycles, epigenetics, and nuclear position) can have different motion behaviors. **(b)** RNA single molecules get through the nuclear pore complex. The transcripts leave the transcription sites and diffuse through the pathways intertwined between compact chromatin domains and compartments until the nuclear pore complex. **(c)** CTCF/YY1 and cohesin anchor the chromatin loop. **(d)** The dynamics of nucleosomes and chromatin fibers. **(e)** Transcription factors (TFs) search the genome by 3D diffusion and one-dimensional sliding with fast kinetics until the target site in the enhancer region is encountered. **(f)** Epigenetic effectors dynamically interact with chromatin. **(g)** The chromatin remodeling process. **(h)** DNA repair machinery. **(i)** Telomeres form T-loops, which consist of complicated ribonucleoprotein. (source: S. Shao, Xue, and Y. Sun 2018)

4.1.3 Alternatives to SPT to image transcription

SPT is not the only imaging technique that can provide information about transcription. For completeness, we present a few techniques alternative to SPT.

1. **Imaging single locus.** Despite its attractiveness, most SPT techniques do not provide sequence specificity: it is not possible to know on which sequence a specific DNA-binding event happens. Several techniques were developed to image specifically a given locus.

   (a) *In fixed cells.* One of the first techniques to image transcription is fluorescence in-situ hybridization (FISH), that uses a DNA probe complementary to the sequence of the locus to be imaged. FISH has been of tremendous use to image transcription, but is restricted to fixed cells. In the recent years, FISH has been scaled up so that hundred of loci can be imaged in a single cell (Fields et al. 2018; G. Nir et al. 2018; Mateo et al. 2019).

   (b) *In live cells, using Cas9 imaging.* Intense efforts have been devoted to translate locus imaging to live cells. To do so, the RNA-guided Cas9 protein is fused to a fluorescent protein, and a specific gRNA complementary to the target sequence is transfected into cells. Imaging and tracking of a single locus has been shown using this technique (B. Chen et al. 2013; Qin et al. 2017; Martens et al. 2018).

   (c) *Imaging genomic regions.* When one focuses on a given nuclear compartment, such as active genes, other labeling techniques have been used, that trade the single-gene specificity for a genome-wide labeling. For instance, the ATAC-see technique labels all accessible regions, that can then be imaged, and further correlated with other labelings (X. Chen et al. 2016). ATAC-see uses the property of the Tn5 transposase, that only integrates at accessible regions of the genome. Similarly, active gene compartments can also be visualized in live by looking at the incorporation of labeled modified nucleotides (D. K. Sinha et al. 2008; W. Xiang et al. 2018).

2. **Imaging single mRNAs.**

   (a) *In fixed cells.* Similar to DNA FISH, RNA FISH has been developed to image single mRNAs in cells. Similar to DNA FISH, the technique was scaled up to image up to 10,000 mRNA species in parallel using a multiplex FISH technique (Eng et al. 2019). Similar to DNA FISH, RNA FISH can only be applied to fixed cells.

   (b) *In live cells.* Several techniques have been developed to image single mRNAs in live cells. The first one is the MS2 and MS2-derived system: a series of MS2 stem loops are added to the DNA sequence in 5' or 3' of the gene. When transcribed into RNA, the stem loops can be specifically bound by a protein, fused to a fluorescent protein. This technique allows to investigate the dynamics of single mRNA molecules in live cells.

   Other techniques avoided the editing of the gene by relying on the complementarity between RNA and other fluorescently labeled species, such as RNAi (Avivi et al. 2017) or morpholinos (Hadzhiev et al. 2018) and provided a live-cell view of mRNA dynamics.

3. **Imaging proteins.** Several techniques have been developed to image proteins, without the need to image individual molecules. We briefly review the main existing techniques in this section.

   (a) *Fluorescence imaging after photobleaching (FRAP).* FRAP is probably the most widely known live-cell imaging technique. In FRAP, the cell expresses labeled fusion proteins, and a high power laser is used to quickly bleach a defined region of the cell. Due to the combined effect of diffusion and binding/unbinding of the fluorescent proteins, the bleached region progressively "recovers" and becomes fluorescent again (Phair and Misteli 2001). The recovery of fluorescence within the bleached spot is monitored over time, and the speed and

shape of this recovery provides information on the diffusion of the protein of interest. It is then possible to perform modeling and model fitting in order to extract the number of diffusing species and their diffusion coefficients. A vast literature exists on the estimation of diffusion parameters from FRAP (Periasamy and A S Verkman 1998; Phair, Gorski, and Misteli 2003; Mueller, Davide Mazza, et al. 2010; Daddysman and Fecko 2013; Lorén et al. 2015).

(b) *Fluorescence correlation spectroscopy (FCS).* In FCS, the fluorescence intensity inside a confocal volume is tracked across time: it increases when one fluorescent molecule enters the confocal volume and decreases when it exits it. Fast-diffusing molecules cross the confocal volume faster than the slow-diffusing ones, yielding a detectable signature. Despite being a conceptually simple technique, there is a proficient literature on how to analyze FCS data. Most of the analysis focused on computing aggregated statistics such as auto-correlation curves followed by model fitting (McHale, Berglund, and Mabuchi 2004; S.-M. Guo et al. 2012; J. He, S.-M. Guo, and Bathe 2012; G. Sun et al. 2015; Tsekouras et al. 2015). More recent development use the full temporal information: the timestamp of photon arrivals at the detector and require much shorter acquisition times (Jazani et al. 2018), simultaneously quantify diffusion and anomalous diffusion (F. Schneider et al. 2018) or combine it with SPT (Limouse et al. 2017).

(c) *Assessing protein-protein interactions: Förster resonance energy transfer (FRET).* In FRET, two fluorophores with compatible spectra are used: one "donor" fluorophore on one protein or one protein domain, and one "acceptor" on another protein or the other protein domain. When the two fluorophore are close enough ($<$ 1nm), excitation of the low wavelength fluorophore yields a non-radiative energy transfer towards the other fluorophore, that starts emitting light, despite not being directly excited. As such, the ratio of fluorescence between the two colors provide a quantitative estimate (Hellenkamp et al. 2018) of the distance between the two fluorophore, and indirectly of protein-protein interactions (Margineanu et al. 2016). FRET is widely used *in vitro* and has been repeatedly used to monitor histone modifications dynamics in live cells (Sasaki et al. 2009; Peng et al. 2018).

4.2 Introduction to diffusion

Diffusion is the physical process by which atoms, molecules and small objects move under thermal agitation. It is unclear when it was first observed, but some modern Latin poems mention the diffusion of dust (Lucretius's scientific poem "On the Nature of Things"; c. 60 BC). Despite an early description, a mathematical formulation of diffusion has been lacking for long, and was only proposed in the 18th and 19th century.

In this section, we introduce some mathematical background about diffusion and diffusion processes. This chapter might be a little bit technical but can be skipped without problem. We present here a mathematical formulation of diffusion, and derive some useful characterizations, such as the mean squared displacement.

4.2.1 Free diffusion and Brownian motion

Free diffusion (or Brownian motion) can be derived using various approaches. Indeed Brownian motion connects the fields of continuous time and discrete time analysis, and of stochastic and partial differential equations. We present here three complementary perspectives on Brownian motion (Grebenkov 2009; Briane, Vimond, and Kervrann 2018):

1. **Partial Differential Equation (PDE) formalism.** Brownian motion can first be derived from Fick's law, following an approach developed by Einstein in 1905. Fick's law relates the flow of particles to the concentration gradient and states that the flow is proportional to the gradient, with a proportionality constant D, the diffusion coefficient:

$$J = -D\nabla C,$$

with J the flow (in units of particles per surface area per time unit; $mol.m^{-2}.s^{-1}$), D the diffusion coefficient (in units of surface area per time unit; m^2s^{-1}) and C the concentration of diffusing particles (in $mol.m^{-3}$).

Then, it is easy to derive the diffusion equation, analogous to the heat equation:

$$\frac{\partial C(x,t)}{\partial t} = -D\Delta C(x,t), \tag{4.1}$$

Where Δ is the Laplace operator : in 1D, the second spatial derivative, $\Delta = \partial^2/\partial x^2$. The diffusion equation thus relates the spatial and temporal evolution of particles.

As with any partial differential equation (PDE), one can find a solution when boundary conditions and initial conditions are specified. For diffusion equations, the solution corresponding to an initial condition where a point source of particles (a Dirac, $\delta(x)$) is placed at position x bears a particular name: it is called the fundamental solution, the propagator or the Green function.

The Green function is a key solution of such family of equation for two reasons: (1) if the Green function of a PDE can be determined, then the solution of the PDE with arbitrary initial condition is known through a simple convolution and (2) for the special case of single-molecule considerations, the propagator describes the motion of an individual diffusing particle.

One can easily check that the following function is a solution to equation 4.1:

$$p(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{\frac{-x^2}{4Dt}} \tag{4.2}$$

This perspective benefits from the very strong power of PDE analysis, and many computations can be performed easily and analytically. However, the PDE formalism does not give access to single trajectories, and only presents average concentrations. These limitations are addressed in the next two subsections.

2. **Discrete time formalism.** A single-trajectory perspective can be intuited from the propagator of the random walk. One can make the analogy between a point-like source of molecules (a Dirac, $\delta(x)$) and a single-molecule. Then, the evolution of the concentration over time, properly normalized, can be seen as a probability of presence over time. Thus, the propagator can be interpreted as the probability density of a particle moving from the origin to a given point.

   Thus, the displacement of a particle between time t and $t + t$ can be drawn from the following probability distribution p:

   $$p(x, dt) = \frac{1}{\sqrt{4\pi Ddt}} e^{\frac{-x^2}{2Ddt}},$$

   that is, the displacements X follow a centered normal law with variance $2Ddt$: $X \sim \mathcal{N}(0, 2Ddt)$.

   This discrete formalism provides a straightforward approach to simulate single trajectories, in a step-by-step manner.

3. **Stochastic Differential Equation (SDE) formalism.** Finally, one can formally take the limit of infinitely small time steps, to obtain a (stochastic) function that describe the continuous motion of a particle. This approach is called a "diffusion approximation". It allows to rigorously describe the motion of a particle for arbitrary time steps. Brownian motion can be written as the following stochastic differential equation (SDE):

$$dX_t = \sqrt{2D}dB_t,$$

where B_t represents a normal Brownian motion ($B_{t-s} \sim \mathcal{N}(0, t-s)$).

4.2.2 Characterization of diffusions

After introducing various formalisms related to free diffusion, we provide two important characterizations of diffusions, and provide results for the case of free diffusion.

1. **Mean Squared Displacement (MSD).** First, diffusion can be characterized by its mean squared displacement (MSD). The MSD describes how a particle explores space over time. It is defined as (assuming a discrete trajectory of N points):

$$MSD(t) = \frac{1}{N}\sum_{n=1}^{N}\left(x_n(t) - x_n(0)\right)^2, \text{ for the continuous case}$$

$$MSD(t) = \mathbb{E}\left(||X_t - X_0||^2\right), \text{ for the SDE case.}$$

Under free diffusion, using the fact that the propagator follows a normal law of variance $2Dt$ one can easily compute that:

$$MSD_{normal}(t) = 2Dt.$$

Thus, computing the slope of the MSD allows to get an estimate of the diffusion coefficient. When the diffusive process observed deviates from free diffusion, a phenomenon called anomalous diffusion, the MSD can no longer be fitted by a straight line, but often by a power law:

$$MSD_{anomalous}(t) \sim t^{\alpha},$$

with α the anomalous diffusion exponent, between 0 and 1. When $\alpha = 1$, one recovers normal diffusion. The computation of the MSD is illustrated in Figure 4.2-left.

2. **Jump length distribution.** A second characterization of diffusion can be derived by looking at the *distribution* of jump lengths, rather than the *mean* squared displacement, thus providing a richer information. The jump length distribution gives the probability for a particle to make a jump of given length r as function of the time lag (Δt). For free diffusion, it can be derived by radially integrating the propagator:

$$p(r, \Delta t) = \frac{r}{2D\Delta t}e^{-\frac{r^2}{4D\Delta t}}.$$

Again, this formula provides another manner to estimate the diffusion coefficient, by performing for instance a non-linear least squares fit. This is the strategy adopted by the Spot-On tool, presented in section 1.3. It is illustrated in Figure 4.2-right.

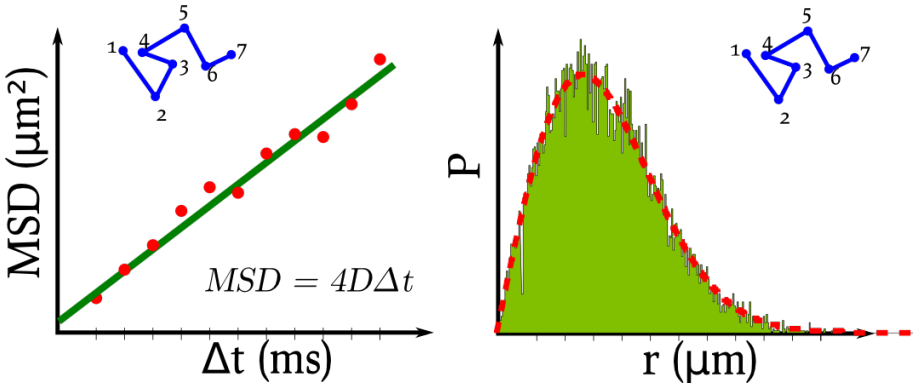


Figure 4.2: Computation of the MSD (left) and the jump length displacement histogram (right).

4.3 Difficulties in analyzing SPT data

In the previous section, we presented three mathematical descriptions of Brownian motion. In this section, we highlight some of the frequent deviations from ideal diffusion processes that are usually observed, and that greatly complicate the analysis of SPT data. These biases need to be taken into account when designing a SPT analysis algorithm (such as the ones detailed in section 1.3).

4.3.1 Noise

First of all, when single molecules are imaged under a fluorescence microscope, a localization error, σ, is associated with each molecule, that depends, on a first approximation, on the number of photons collected and the background level (Mortensen et al. 2010).

In practice, the true position of the particle is corrupted by an (assumed) normal Gaussian noise of standard deviation σ. Fortunately, the two metrics presented above can be easily adjusted to incorporate localization error, under a free diffusion assumption. The MSD is given by:

$$MSD^{\sigma}_{free}(t) = 4\sigma + 2Dt.$$

Note that this estimator is usually regarded as a poor estimator, and many criticism has been voiced against the use of MSD to characterize protein motion. Indeed, at low time intervals, the estimate of the MSD is contaminated by localization error. At higher intervals, little data is available. Thus the range where data can be used is narrow (Michalet 2010). More detailed calculations accounting for noise were performed in (Deschout, Neyts, and Braeckmans 2012).

The jump length displacement is given by:

$$p^{\sigma}_{free}(r, \Delta t) = \frac{r}{2(D\Delta t + \sigma^2)} e^{-\frac{r^2}{4(D\Delta t + \sigma^2)}}.$$

4.3.2 Confinement

Second, these two formulas are valid when no boundary conditions are assumed. When a particle diffuses inside a cell, it is confined within the cell membrane. Confinement yields non-linear MSD, that saturates for high times considered. The MSD of a particle confined in a 3D sphere of radius a can be computed as (Bickel 2007):

$$MSD(t) = \frac{6a^2}{5} - 12a^2 \sum_{n=1}^{\infty} \exp\left[-\beta_{1_n}^2 \frac{Dt}{a^2}\right] \frac{1}{\beta_{1_n}^2(\beta_{1_n}^2 - 2)},$$

with β_{1_n} the (non-zero) zeros of the derivatives of the spherical Bessel functions, $j_l'(\beta_{l_n}) = 0$. One can see that at long time scales ($t >> 1$), the MSD reaches a plateau: $\lim_{t\to\infty} MSD(t) = a^2$.

4.3.3 Motion blur

When a particle in motion is imaged under a constant exposure time, it appears blurred: the photons emitted by the diffusing particle are spread out. Motion blur has several consequences. Because photons are spread out, localization accuracy is reduced and most fitting algorithms might miss the detection. Furthermore, many metrics to quantify diffusion are biased when motion blur is present (Berglund 2010). Thus, a rigorous treatment of motion blur (either at the experimental or the analysis level) is needed. In section 1.3, we present an experimental approach to avoid motion blur. Conversely, the detection technique presented in section 1.2 includes motion blur, that will have to be taken into account at the analysis level.

4.3.4 Trajectory length

Third, the length of trajectories that can be observed using SPT is limited by two factors. The lifetime of the fluorophore under the high laser exposures required for high signal-to-noise ratio leads to short trajectories. Moreover, particles tend to exit the focal plane as they diffuse, thus limiting the length of the observed trajectories (Matsuoka, Shibata, and Ueda 2009; Kues and Kubitscheck 2002). This effect is exemplified in Figure 4.3.

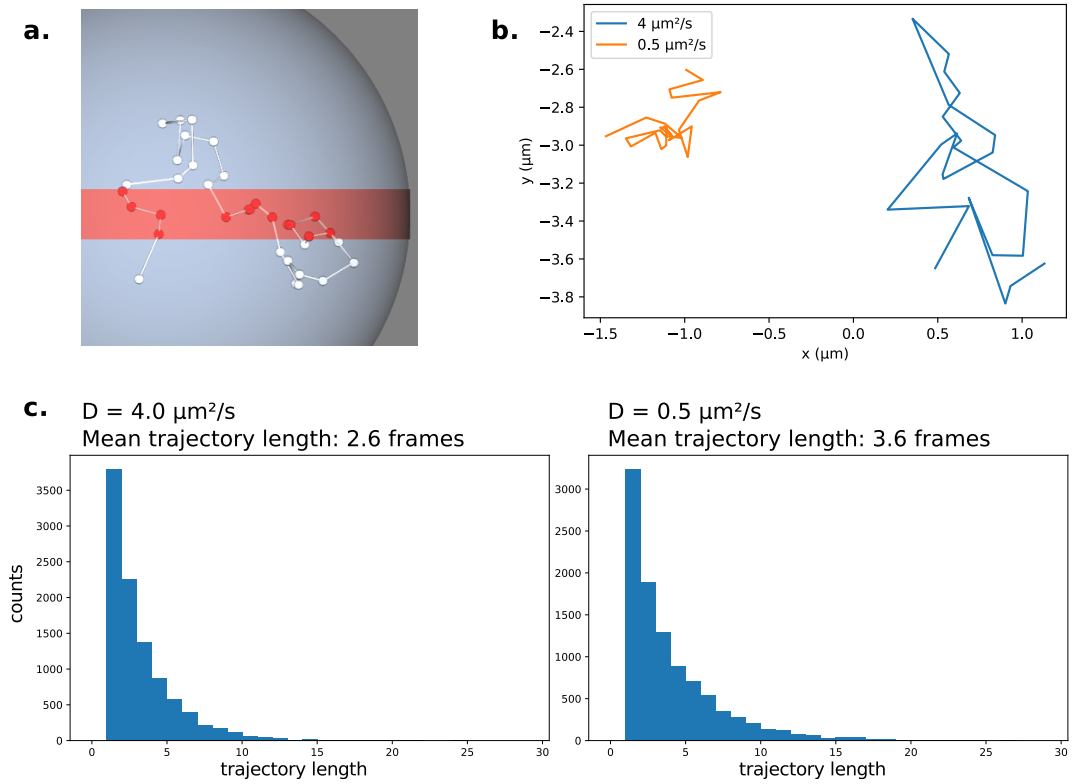


Figure 4.3: **Empirical limits on the trajectory length. (a)** Because the axial detection range of the objective is limited, diffusing particles move in and out of focus, leading to artificially short trajectories. **(b)** Sample trajectories simulated with $D = 0.5\mu m^2/s$ and $D = 4\mu m^2/s$, note the longer displacements for the fast diffusion. **(c)** Histogram of trajectory length for two sets of simulations. Fast-diffusing particles yield significantly shorter trajectories (mean length of 2.6 frames vs. 3.6 frames).

4.4 Critical review of existing SPT analysis techniques

4.4.1 Slow SPT analysis techniques

1. **Canonical slow SPT analysis.** In slow SPT, only the particles bound to DNA are visible. By tracking over time bound particles, one can estimate the residence time of the protein on DNA. This is usually done by computing the distribution of residence time from multiple tracks (that is, the distribution of track lengths), and performing a model fit to extract one or several residence times.

   Briefly, and following (Anders S. Hansen, Pustova, et al. 2017), the residence time on chromatin is usually assumed to follow a first order kinetic: the time spend on chromatin (t_{ON}) then follows an exponential distribution of parameter $k = 1/t_{ON}$, and a constant dissociation rate can be estimated. Under this simple hypothesis, the time spent on chromatin for a single binding event follows the following distribution $P(t)$, also termed the survival probability:

   $$P(t) = Ae^{-kt},$$

   with A a normalization constant. This model is too simplistic, however, and needs to be refined, for instance to incorporate two biases: (1) photobleaching can happen a rate k_{pb} and (2) truly bound molecules can occur as a mixture of populations, for instance some molecules binding at specific (S) sites, and some binding at non-specific site (NS). Under this condition, the survival probability can be expressed as:

   $$P(t) = Ae^{-(k_{ns}+k_{pb})t} + Be^{-(k_s+k_{pb})t}.$$

   Often, $k_{ns} >> k_{pb}$ and k_{pb} can be neglected in this term of the equation. Additional biases can be taken into account, such as the slow defocusing of the nucleus over time (Anders S. Hansen, Pustova, et al. 2017).

2. **Other analysis techniques.** Under this model, it is very difficult to disentangle k_{pb} and k_s because they have the same magnitudes. Some approaches were developed in order to get more precise estimates of k_s. One of them replaces the long exposure times (2 Hz) by short frames interspersed by long dark times. By varying the dark time, it is possible to decouple photobleaching from the specific binding rate (Presman et al. 2017; Ho et al. 2019). Such approach has been extended to estimate any number of states (Reisser and J. Christof M. Gebhardt 2017).

4.4.2 Fast SPT analysis techniques

The analysis of SPT has been intensively investigated, and one can distinguish several families of techniques (see also for reviews: Metzler, Tejedor, et al. 2009; Récamier 2013; Ernst, Köhler, and M. Weiss 2014). In the field of stochastic processes, the inference of a diffusion coefficient from a sampled process is a common problem (see for instance Florens-Zmirou 1993; Hoffmann 2001). However, this theory cannot be applied when moving to experimental trajectories, and other approaches have been proposed.

1. **MSD-based techniques.** A first family of SPT analysis algorithms tries to perform robust MSD inference. The use of MSD to study diffusion was introduced by Einstein in 1906, and was revived in biology by (Qian 2000). MSD analysis can either be performed by inferring a diffusion coefficient from a single trajectory (a setting studied in Michalet and Berglund 2012) or by pooling various trajectories (Z. Liu, Legant, et al. 2014), and many refinements and estimators based on the MSD have been proposed (Michalet 2010; Boyer, Dean, Mejía-Monasterio, et al. 2012).

When inferring kinetic parameters from a series of single trajectories, one faces the issue that for common trajectory lengths obtained in nuclear SPT (length of $<<$ 20 points per track) and common localization error, inaccuracy reach 100% (Michalet and Berglund 2012; Fig 1 and Anders S. Hansen, Woringer, et al. 2018). As such, any approach that uses MSD on short trajectories should be evaluated with a lot of suspicion. For longer trajectories (such as diffusion in a membrane), approaches have been proposed that can segment trajectories based on the type of motion (Monnier, S.-M. Guo, et al. 2012).

2. **Hidden Markov Models (HMMs).** A second family of SPT analysis algorithms derived from Markov models and Hidden Markov Models. Most of them were derived to perform trajectory segment classification, the hidden value inferred being the state of diffusion, or the current diffusion coefficient. For instance, (Monnier, Barry, et al. 2015) introduces the HMM-Bayes technique to infer whether a trajectory segment is in one (or several) diffusive or active transport states. Moreover, (Paddy J. Slator, Cairo, and N. J. Burroughs 2015) implemented the inference of localization noise to infer switches in diffusion coefficient within one trajectory. A similar approach was used to detect confinement (Paddy John Slator and N. Burroughs 2018).

   These methods often rely on a fixed number of states, which comes from significant mathematical limitations. Some of these limitations were overcome using so-called variational Bayesian inference (Blei, Kucukelbir, and McAuliffe 2016). The prototypical algorithm performing variational Bayesian inference on a HMM is `vbSPT` (Persson et al. 2013). This algorithm can estimate the number of diffusive states and progressively consolidate increasing information about these states as trajectories are analyzed. The algorithm was further refined to incorporate the estimate of localization error (Lindén and Johan Elf 2018).

3. **Inferring maps of diffusion coefficients.** A third family of SPT analysis algorithms not only infers the diffusion coefficient over the population of diffusing molecules, but also a spatial map of diffusivity (J.-B. Masson et al. 2009; El Beheiry, Dahan, and Jean-Baptiste Masson 2015). This approach has been pioneered in membranes, where a high density of tracks can easily be obtained. Inside cells, these promising techniques have not been tested, but the high diffusion coefficients of nuclear proteins might render such a map difficult. Moreover, unlike in membranes, proteins can reside at the same location with different diffusion coefficient, depending on whether they are interacting with a given structure or not.

4. **Inferring anomalous diffusion.** Fourth, many approaches have been proposed to infer anomalous diffusion in cells. Some of them are reviewed in (Guigas and M. Weiss 2008). A direct technique can be used by fitting the MSD with a power law to estimate the anomalous diffusion coefficient α. However, alternative techniques have been proposed. Many of them focus on the inference of model-specific parameters, or on techniques to distinguish between types of anomalous diffusion.

   (a) *Inferring diffusion parameters for anomalous diffusion.* Several methods have been proposed to infer diffusion parameters for several anomalous diffusion models. For the case of diffusion in disordered (fractal) media, (Shkilev 2018) propose estimators that can be applied to SPT, FCS and FRAP.

   For the case of fractional Brownian motion, techniques to infer both the anomalous diffusion coefficient (α) and the generalized diffusion coefficient (D_α) have been proposed. The former approach (Krog et al. 2018) takes into account noise (localization error) and drift. It uses Bayesian inference. The latter (Boyer, Dean, Mejia-Monasterio, et al. 2013) relies on squared displacements and uses least squares to estimate D_α.

   (b) *Distinguishing between anomalous diffusion models.* While the previous section focused on identifying model parameters, assuming that the type of diffusive process is known, other authors tried to infer the type of diffusive process itself. They try to distinguish between

various anomalous diffusion models. A prototypical approach (Robson, Burrage, and Leake 2012) used Bayesian inference to distinguish between Brownian, anomalous, confined and directed diffusion, and uses the propagators associated with each different diffusion model. Getting more into details, (Hellmann et al. 2011) found using a simulation study that it is very hard to distinguish between fBm and obstructed diffusion when localization noise is present, both in SPT and FCS. The authors used a combination of techniques for the inference, including MSD and p-variation techniques. In (Burnecki et al. 2012), the authors propose a series of test to "unambiguously" identify fBm, by progressively proving that several other models are wrong. Other tests were proposed to distinguish fBm from a CTRW (Magdziarz et al. 2009) using a test based on p-variations. The p-variations are the finite sum of the p-th powers of the increments of the trajectory. Finally, approaches inferring the mean first passage time of a particle were used to distinguish between CTRW and diffusion in fractals (Condamin, Bénichou, et al. 2007; Condamin, Tejedor, et al. 2008).

5. **Other approaches.** Finally, many other families of techniques were proposed. Some relied on maximum likelihood estimates (Thapa et al. 2018), auto-correlation functions (S. C. Weber et al. 2012) or on more exotic estimators (Vestergaard, Blainey, and Flyvbjerg 2014).

   Another line of progress was made in the type of models being simulated. For instance, (Amitai 2018) introduced a model in which TFs can bind and rebind in a dense chromatin mesh. This model was successively fitted to explain anomalous diffusion of CTCF dynamics (Anders S. Hansen, Amitai, et al. 2018).

   Finally, we note that many models were developed to infer trapping potential in membranes (Türkcan, Alexandrou, and Jean-Baptiste Masson 2012; Jean-Baptiste Masson et al. 2014 for instance). We do not review them here since their application seems limited to membranes.

4.4.3 Techniques from locus tracking

Some of the models presented above were never used in the challenging context of nuclear SPT. It is likely that many will not perform well under the conventional imaging conditions of the nucleus. However, the diffusion of brighter and slower-diffusing objects such as DNA loci has been investigated. We mention them here as first examples of experiments bridging the gap between some inference models and real-life experiments.

In addition to anomalous diffusion and its heterogeneity, that can easily be inferred from locus tracking (for instance for telomeres in Bronstein et al. 2009), it has been shown that more specific metrics such as the mean first passage time and the type of diffusion can also be inferred, for instance in the case of the VDJ recombination in B-cells, a key process in the production of antibodies in response to an infection (Lucas et al. 2014). Moreover, inference based on more complex chromatin models, for instance deriving from a spring-and-binder model have been proposed to explain the dynamics of loci, for instance when a DNA break occurs (Amitai et al. 2017). Finally, specific analysis techniques such as two-loci tracking have been developed for the case of DNA tracking (Shukron and Holcman 2017)

4.4.4 Perspectives

We conclude this section by a discussion on SPT analysis techniques.

1. **Seeing anomalous diffusion.** A first comment is the extent by which anomalous diffusion really exists in cells for diffusing proteins. This question was raised by Saxton in one of its last paper published (Saxton 2012). Indeed, despite decades of research, clear benchmarks of anomalous diffusion are missing. Non-linear MSDs have clearly been observed repeatedly. However, whether the anomalous diffusion observed is caused by transient or permanent anomalous diffusion is not clear, and the range of anomalous diffusion has not been fully investigated.

2. **Towards a challenge for a critical evaluation of those techniques.** To do so, in addition to high-quality benchmark materials, good quality analysis algorithms are needed. So far, only limited comparisons have been performed between algorithms (see for instance Weimann et al. 2013; Anders S. Hansen, Woringer, et al. 2018) and a comprehensive benchmark of techniques is missing, despite some framework being proposed (Rigano, Galli, et al. 2018). We propose a collaborative challenge to tackle this issue in section II.4.

3. **What do we care about?** Finally, as highlighted in section 2, it is often possible to gain a lot of information on the diffusive process without getting a fine characterization of the type of diffusion. This idea was put forward in (Bénichou et al. 2010), and suggests that a key parameter to estimate is whether the random walk is doing a compact, recurrent walk, or a non-compact, transient walk, since this parameter determines most of the useful properties of the particle.

Part II

Results

Chapter 1

Developing tools to analyze SPT

Introduction. The main part of my PhD was to develop tools to analyze single-particle tracking (SPT) data, with a focus on SPT data acquired in the nucleus of live cells.

In order to be efficient, a SPT analysis tool should satisfy several properties: (1) *performance*: it should be good at estimating diffusion parameters, (2) *robustness*: it should be able to provide good estimates, despite the limitations and biases of the acquisition system, (3) *ease of use*: a non-programmer user should be able to use it and understand the basic principles of the program. Moreover, the tool should be widely available.

In practice, such a tool has then to: (1) be designed with many of the biases of SPT in mind. (2) be validated against a wide range of simulated and experimental conditions. (3) be available as an open-source tool and sufficiently documented.

In this section, we will review the analysis techniques that we developed. First, we designed a tool, simSPT (Anders S. Hansen, Woringer, et al. 2018; section 1.1, also used for the challenge II.4), to simulate very quickly realistic SPT data. Then, we preliminary explored the possibility to perform 3D SPT, rather than 2D (section 1.2). Finally, we explored various options to analyze SPT data:

- an approach based on the modeling of the jump length distribution, inspired from D. Mazza et al. 2012 (section 1.3)
- a refinement of the previous approach, that takes into account anomalous diffusion processes (section IV.2)
- an approach based on the joint modeling of the jump length distribution and the distribution of angles (section IV.3)

1.1 Development of a simulation tool

Introduction. During the development of any inference algorithm, one should perform benchmarks against a "ground truth" (a series of measurements in which the "true" value is known). In the case of SPT, this ground truth can be obtained from experiments using calibrated solutions of known viscosity in which particles of known size diffuse. Another way of getting a ground truth is to perform SPT simulations *in silico*.

Thus, in order to validate the data analysis tools that we produced, we wrote a fast simulation tool, simSPT. This tool tries to balance three variables: simulation speed, ease of use and realism of the simulations. simSPT can simulate both Brownian and non-Brownian motion of proteins confined in a sphere or in a cube, and takes into account several photo-physical parameters such as the photo-bleaching rate and the number of photons emitted, through the localization error. In this section, we review the rationale and implementation of simSPT.

1.1.1 The simSPT framework

Despite the fact that Brownian motion can be trivially simulated in a few lines of code (section 4.2), several significant barriers need to be overcome in order to make the simulation realistic.

Several approaches can be undertaken in order to simulate SPT. One one side, simulations based on molecular dynamics (MD) with explicit solvent have no limitation in terms of the type of motion they can simulate, the choice of the confinement geometry, etc. MD simulation, however, are usually difficult to parametrize and time consuming. On an intermediate scale, some tools simulate (Lindén, Ćurić, Boucharin, et al. 2016) TIFF images of Brownian motion or trajectories (Drawert et al. 2016).

SMeagol by (Lindén, Ćurić, Boucharin, et al. 2016) is one of the most advanced simulation tools. It simulates "3D diffusion in cellular compartments, diffusion limited reaction kinetics, surface adsorption, reactions in membranes and other complex aspects of reaction diffusion kinetics that do occur in cells, but are not considered in SPT analysis algorithms". The tool also integrates the simulation of a 3D PSF, the kinetics of photo-activation, blinking and bleaching of the simulated fluorophores, background noise and camera specific parameters. Despite being highly versatile, the code of SMeagol suffers a few limitations. First of all, it is limited in terms of speed. It requires that many parameters are provided by the user, and finally, it is limited to the simulation of Brownian motion. `simSPT` can be seen as a simpler alternative to SMeagol.

`simSPT` is a small tool to simulate single particle trajectories of freely diffusing molecules in a confined geometry (so far, a cube or a sphere) observed in fluorescence microscopy. `simSPT` tries to be realistic about that, that is it takes into account (uneven) photobleaching (simulating fluorophore lifetime and a HiLo beam; Tokunaga, Imamoto, and Sakata-Sogawa 2008), it also takes into account a user-defined detection probability along the z axis. Finally, `simSPT` can simulate state transitions (using the Gillespie algorithm). In all cases, simulations are performed starting from a steady state. Furthermore, in its latest version, `simSPT` can handle several models of non-Brownian diffusion. Also, `simSPT` is written in C, and is fast (that is, >> 20 times faster than the corresponding Python/Matlab implementation).

We are very thankful to Anders S. Hansen who designed the initial simulation scheme, for comments and insightful suggestions. `simSPT` incorporates `read.c` from Scott Brueckner (Sept. 1999) and fractional Brownian motion simulation routines by Ton Dieker (Dieker 2004). This section borrows some text from the simSPT documentation.

1.1.2 Simulation of realistic Brownian motion

1. **Brownian motion & photophysics** As explained in section 4.2, Brownian motion, that is the trajectory over time of one freely-diffusing particle, can be simulated by drawing normal random numbers with mean 0 and standard deviation $\sqrt{2Ddt}$ where D is the diffusion coefficient of the particle and dt the exposure time. Indeed, for Brownian motion, the position X_{t+dt} of the particle at time $t + dt$ can be computed as $X_{t+dt} = X_t + \sqrt{2Ddt}dB_t$, with dB_t a unit, normal random number ($dB_t \sim \mathcal{N}(0, 1)$). This approach is called the Euler-Maruyama scheme[1].

   When several sub-populations of particles should be simulated, `simSPT` successively simulates the various individual populations, unless the option to account for state changes has been selected (section 3). In the former case, populations are defined by a diffusion coefficient (D), a proportion (p), a fluorophore lifetime (β). In the later case, the tool uses a modified Gillespie algorithm to simulate state changes.

   The particles are simulated with a given lifetime: they photobleach when their photon budget is exhausted. Particles are illuminated when they get inside the HiLo beam (HiLo beam is assumed to be parallel to the (x,y) plane, has the shape of a step function). Inside the HiLo beam, particles are progressively bleached with a constant probability per frame.

[1] In practice, we used an updated version of the Mersenne-Twister algorithm (as implemented in the GNU Scientific Library –GSL–) to generate uniform random numbers. Then, to obtain Gaussian random numbers, we use Marsaglia's ziggurat method (Leong et al. 2005). Prior to version 1.6, `simSPT` was using the Marsaglia Polar Method.

This formulation adopts the formalism of *exponential clocks*: every time a particle is created, the software associate a lifetime. This lifetime is an integer number of frames that is drawn from an exponential distribution whose mean lifetime is provided by the user. Then, every frame where the particle lies within the HiLo volume, this lifetime parameter is decremented by one. When it reaches zero, the particles is considered as bleached and is removed from the simulation.

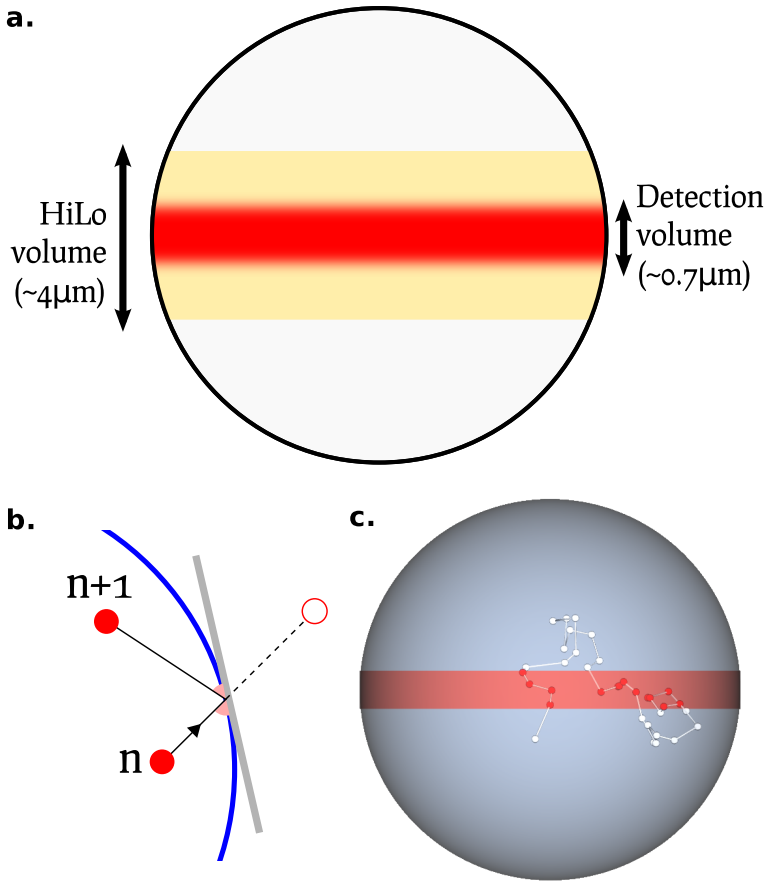


Figure 1.1: **Principle of `simSPT`. (a).** `simSPT` simulates diffusion in a 3D volume (sphere), and diffusing particles are only detected when they are inside the HiLo volume (orange region) and within the depth of field of the objective (red region). **(b).** Principle of specular reflections. **(c).** Example of a simulated trajectory crossing the focal plane.

2. **Confinement and geometry parameters.** Unlike free, unconfined Brownian motion, diffusion in live cells is restricted by the nuclear envelope, or by the cell membrane. It is widely accepted that confinement has a strong influence on the observed trajectories, and thus it should be taken into account.

   More precisely, confinement means that a diffusing particle must remain inside the confining volume (for instance, a sphere representing the nucleus of a cell). From a numerical point of view, this can be achieved by three ways:

   (a) By *resampling*: every time a jump brings a particle outside the confinement volume, this jump is discarded and a new jump is sampled.

   (b) By *substeps*: the whole trajectory is simulated using very fine substeps. For instance, one jump is simulated as the summed displacements of 20 substeps, each representing the motion of the particle over $dt/20$ timesteps. At such a small scale, resampling can easily be performed without introducing major biases.

(c) By *specular reflections*: one assumes that the particle makes "elastic reflections" on the confining surface.

`simSPT` implements the third approach (specular reflections; Figure 1.1b). This choice has been motivated in the literature (Sharp et al. 1987) and many approaches allow a fast computation of such reflections. Conversely, the resampling approach strongly biases many statistics of the random walk, leading to unreasonable distribution of jump lengths. On the other side, simulation using substeps is time consuming: simulating 20 substeps per step is twenty times slower.

In practice, we thus implemented confinement by performing specular reflections on the limits of the simulation volume (see Figure 1.1). The details of the calculations and implementation of 3D specular reflections can be found on the internet.

3. **Usage.** Once compiled (`simSPT` is written in plain C), `simSPT` is a command-line tool. All the parameters are then accessible through command-line arguments, or by providing an input file (only for the case of the simulation of heterogeneous diffusion). Thus, the command line can be used to specify the exposure time (`-dt`), the localization error (`-sigma`), the number of trajectories to simulate (`-n_traj`), the radius of the confinement sphere (`-radius`), the seed of the pseudo random number generator, in order to obtain reproducible simulations (`-seed`), etc. An exhaustive list of the parameters is available in the documentation of the software.

   As an illustration, we present the procedure to simulate a two states model, with a slow fraction diffusing at 0.001 µm²/s, a fast fraction diffusing at 1 µm²/s, and a relative fraction of 50/50%. The localization error (`-sigma`) is set to 35 nm, the frame rate to 10 ms (`-dt`). 100000 trajectories are simulated and stored in `out.csv`. The pseudo-random number generator (PRNG) is initialized at 0 (meaning that this simulations is reproducible, and will always provide the same result). This option can be removed to get a random initialization of the PRNG. An illustration of a sample trajectory is presented in Figure 1.1c.

```
./simSPT -D1=0.001 -D2=1.0 -p1=0.5 -p2=0.5 -sigma=0.035 -dt=0.01 -n_traj=100000
 -file=out.csv -seed=0
```

4. **Performance.** `simSPT` has been designed to simulate data as fast as possible. Benchmarks by Anders Hansen indicate that `simSPT` can generate half a million trajectories in just a few seconds. The entire dataset used in (Anders S. Hansen, Woringer, et al. 2018) is about 80 GB in size and was generated in less than an hour.

1.1.3 Beyond Brownian motion

In addition to the simulation of mixtures of Brownian motion, we implemented the simulation of various types of anomalous diffusion models. The simulation of anomalous diffusion has been the subject of intense research, and each anomalous diffusion model has to be simulated using specific methods (Saxton 2007b).

Most of the existing programs to simulate anomalous diffusion focus on one type of models. Here, we try to implement a diversity of models.

1. **Anomalous diffusion.** So far, we have only considered two simple cases: (1) *one sub-population case*: all the particles follow the same laws of diffusion throughout the entire observation time; Figure 1.2a (2) *multiple sub-population case*: all the particles follow the same laws of diffusion throughout the observation time, but the particles can belong to different sub-populations with different diffusive properties. These two cases can be considered as instances of Brownian motion, since when one follows one given particle through time, this particles displays a (potentially confined) Brownian motion (Figure 1.2b).

However, as explained in section 3.3, many features of the nucleus of a cell can induce a departure from Brownian diffusion. Among them: the existence of regions of space (clusters) where diffusion is modified (section 2), the existence of state changes through time (a particle can switch between states; Figure 1.2c and section 3), diffusion interspersed by long binding times (section 4), diffusion in a viscoelastic medium (section 6) and diffusion in a highly crowded medium (section 5).

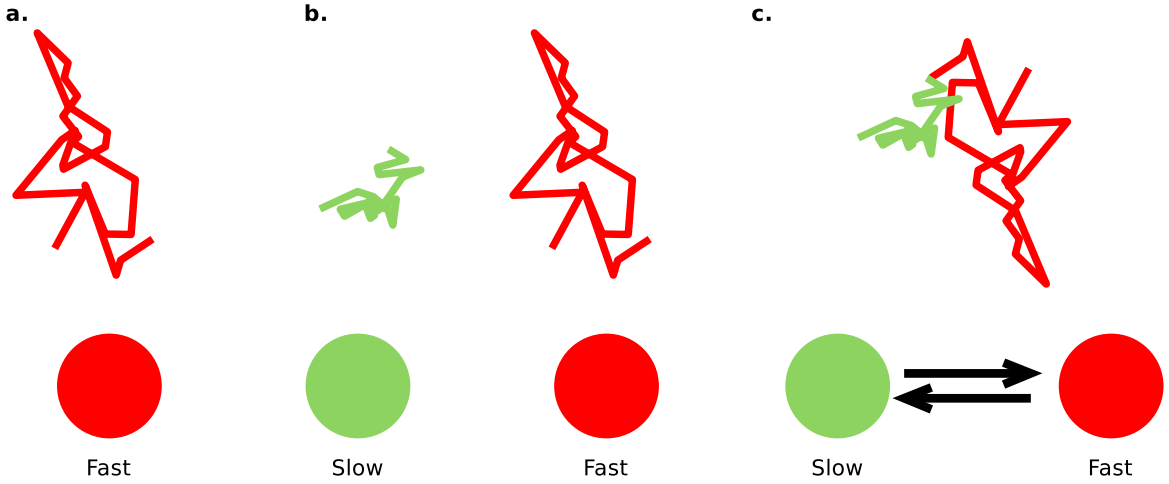


Figure 1.2: **Particles diffusing under a mixture of diffusion coefficients.** **(a)** One population. **(b)** Two populations without state changes. **(c)** Two populations with state changes.

All these phenomena lead to a departure from normal diffusion, namely, anomalous diffusion. Their initial characterization was performed using the mean squared displacement introduced in section 4.2. When diffusion is anomalous, the MSD scales as $MSD \sim t^{\alpha}$ instead of $MSD \sim t$, with $0 < \alpha < 1$. As such, anomalous diffusion is usually characterized solely by the non-linear scaling of the MSD. However, the α exponent, despite being very informative (Bénichou et al. 2010), does not provide mechanistic insights into the process that leads to the observed motion. Thus, more refined characterization of anomalous diffusion are needed. A first step was the implementation of simulation tools, that allowed to simulate anomalous diffusion in the same context as for the Brownian simulation (nuclear confinement, short fluorophore observation duration, etc).

2. **Clusters/heterogeneous diffusion.** The nucleus of a cell is not a homogeneous medium. It can contain zones of fast and slow diffusion, regions of enrichment, etc.

   `simSPT` has basic support for diffusion of a particle in a heterogeneous medium (Figure 1.3). This support so far is limited to the following hypotheses:

   - Depending on which region of space it resides in, the particle has a different fraction bound and enrichment compared to the baseline of the nucleus
   - There is no between regions of space
   - Those regions of space can be specified using a specific text file that lists all the clusters, their location and sizes.

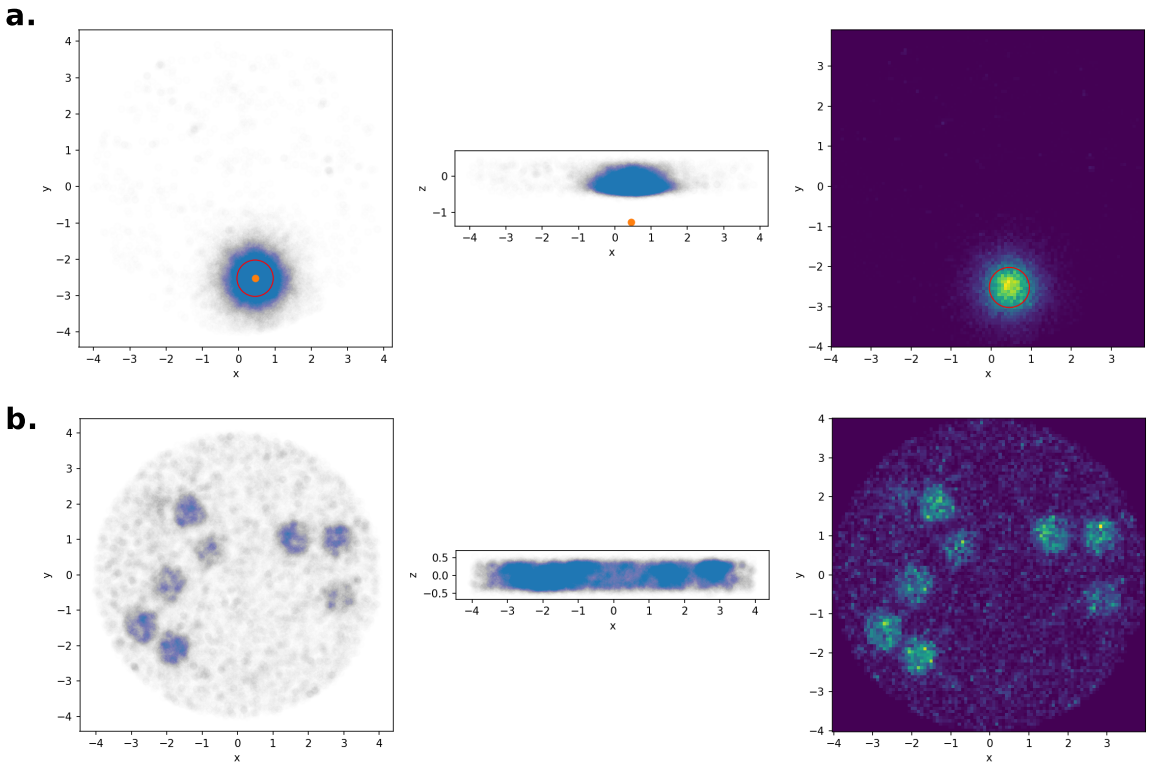


Figure 1.3: **Simulation of clusters of various densities in `simSPT`. (a)** Simulation of one cluster, located slightly under the focal plane (red circle and orange dot), the cluster has a ten times enrichment compared to the rest of the nucleus. **(b)** Simulation of several clusters, under the same conditions as above. *Left.* localizations in the (x,y) plane, *center* localizations in the (x,z) plane, *right* density of localizations in the (x,y) plane.

3. **State changes.** So far, we only presented a simulation setting in which particles of one sub-population always remain in this sub-population. In practice, this is rarely the case and proteins can switch state. For instance, a transcription factor bound to DNA can detach and become freely diffusing. This is traditionally taken into account by assuming single-order kinetics: transition between states a and b can be expressed using the kinetic rates $k_{a\to b}$ and $k_{b\to a}$. When more states are present (a, b, c, representing a fast, slow and bound sub-population for instance), one can define K, a matrix of transition rates, where the element K_{ij} corresponds to the rate $k_{i\to j}$ (Figure 1.2.

   In practice, simulations can be performed using the fact that when first-order kinetics are assumed, the time spent in a state a follows an exponential law of parameter $k_{a\to b}$. When more than two states are considered, the time spent in state b is the minimum of the exponential variables of parameters $k_{a\to b}$ and $k_{a\to c}$, ... This formalism can be easily simulated using a modified Gillespie algorithm. Once initialized at steady-state, the time spent in one state is determined as specified above and particle motion is simulated using the diffusion coefficient of this state. When a state change occurs, the time spent in this new state is also determined and the new diffusion coefficient is used. In case a state change occurs between two frames, we recursively simulate substeps so that the jump length is also computed exactly at this step.

   (a) *Initialization at steady-state.* In order to initialize the simulations, `simSPT` computes the equilibrium fractions ("relative concentrations" at steady-state) achieved under this matrix of transition states. The steady-state distribution is determined by extracting the discrete Markov Chain probability matrix P from the matrix of transition rates K. This is done by normalizing the matrix by the transition rates.

Then, the discrete-time steady state π is determined as the solution of the $\pi P = \pi$ equation (Moler and Van Loan 2003). The continuous-time steady state is finally derived by renormalizing π by the transition rates.

(b) *Simulation of a two-state model with state changes.* As an example, this `simSPT` command-line will simulate two sub-populations with inter-conversion:

- $k_{1\to2} = 25.0s^{-1}$ (`k1_2=25.0`)
- $k_{2\to1} = 25.0s^{-1}$ (`k2_1=25.0`)

```
./simSPT -D1=0.001 -D2=1.0 -k1_2=25.0 -k2_1=25.0 -sigma=0.035 -dt=0.001
 -n_traj=100000 -file=170621_D1_p0.5_dt0.001s.csv -seed=0
```

4. **Continuous time random walk.** A first model of anomalous diffusion is the Continuous Time Random Walk (CTRW). Imagine a scenario where a particle alternates between free diffusion and transient binding (immobilization) as exemplified in Figure 1.2c. If the binding/unbinding rates follow a first-order kinetics (characterized by two kinetic constants, $k_{\text{unbound}\to\text{bound}}$ and $k_{\text{bound}\to\text{unbound}}$ and a time in each state characterized by exponential distributions), then two time regimes can be distinguished:

   - At short time scales, diffusion is similar to the two-states model described above (section 3).
   - At long time scales (observation time much longer than the characteristic time in one state), diffusion and binding times are averaged, and the motion appears Brownian, a consequence of the Central Limit Theorem.

   In the case where the bound time follows a heavy-tailed distribution rather than an exponential distribution, a specific case of CTRW is visible, in which anomalous diffusion is visible. Mathematically (Weissman, G. H. Weiss, and Havlin 1989), a heavy-tailed CTRW can be described by two probability distributions:

   - The (spatial) distribution of jump lengths Φ, that is the same as for Brownian motion, namely $\Phi(x) \sim \mathcal{N}(0, < dr^2 >$
   - The (temporal) distribution of jump times Ψ, that follows heavy tail distribution, such as a power law: $\Psi(t) \simeq \frac{\alpha t_0^\alpha}{\Gamma(1-\alpha)t^{1+\alpha}}$.

   In that case, the time-averaged MSD displays normal, Brownian diffusion (Figure 1.4d-e), whereas the ensemble-averaged MSD displays anomalous diffusion (Figure 1.4b-c; Y. He et al. 2008).

   Following this formulation, it is very easy to adapt the Gillespie simulation algorithm to simulate CTRW, the jump times just have to be drawn from a heavy-tailed distribution (chosen as a power law in `simSPT`) instead of an exponential distribution.

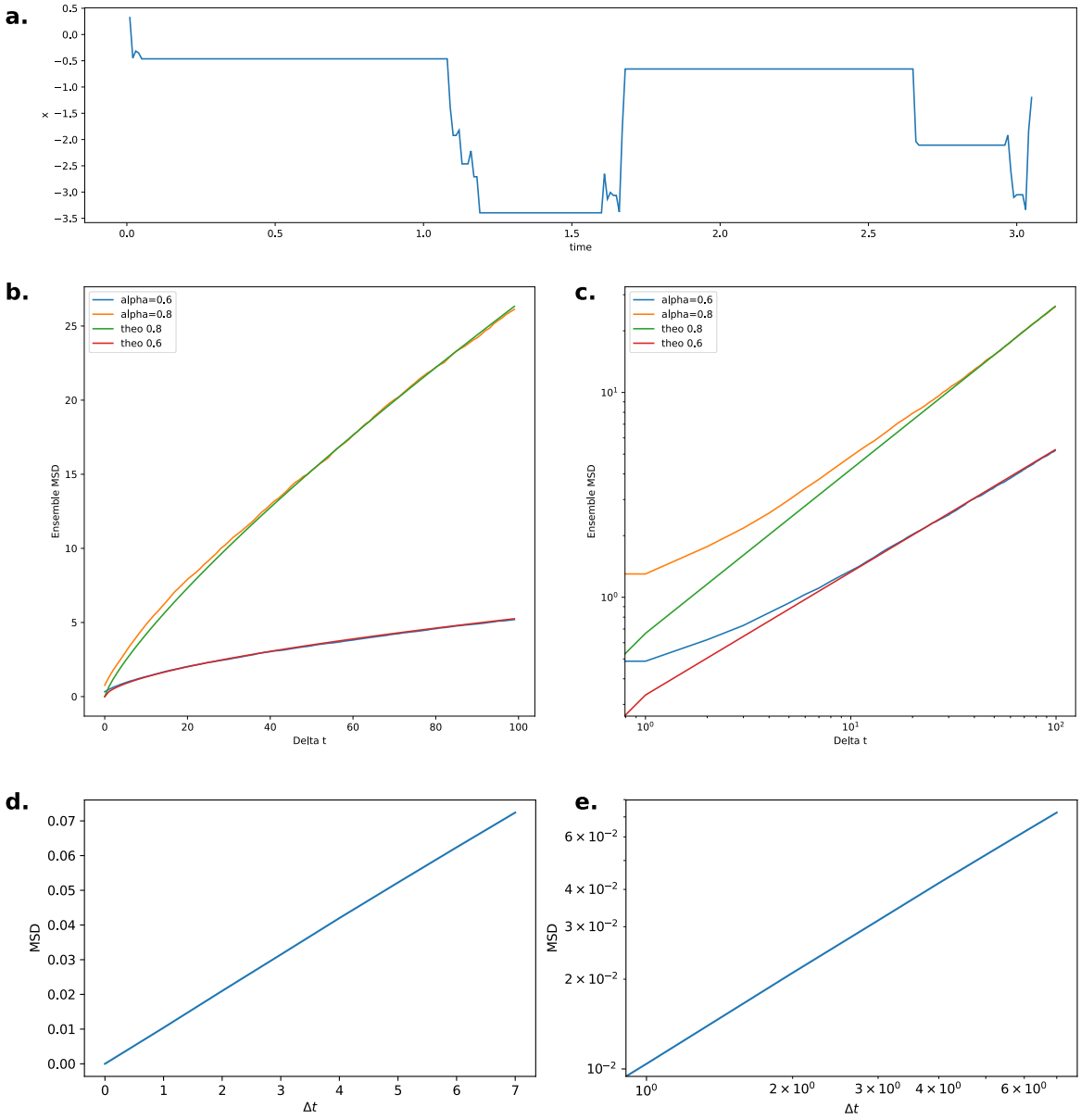


Figure 1.4: **Simulation of CTRW**. **(a)** A 1D trajectory of CTRW, consisting of 250 jumps over 327 time points. **(b)** Comparison of the ensemble averaged MSD between simulated CTRW (*alpha*) and the theoretical MSD (*theo*) for two levels of anomalous diffusion ($\alpha = 0.6$ and $\alpha = 0.8$). **(c)** Same as (b) but in a log-log scale. **(d)** time-averaged MSD for a simulated CTRW of 100000 steps and **(e)** in log-log scale.

5. **Fractals.** Diffusion on fractals is another type of diffusion that yields anomalous diffusion. A fractal is a self-similar structure that reproduces at various scales. When a particle is said to diffuse on a fractal, it is assumed that it diffuses freely on this self-similar structure, and the anomalous diffusion exponent α is observed in the regular Euclidian space. When a particle diffuses in a fractal media, it macroscopically exhibits anomalous diffusion, and the degree of anomalous diffusion depends on a characteristic of the fractal, its fractal dimension d_f.

   In practice, in biology, structures are never self-similar at all scales, because the size of the objects considered are bounded by the size of the nucleus and the size of an atom, thus defining a self-similarity regime. Within this regime of sizes, anomalous diffusion can arise. Despite the fact that many processes can generate fractal structures, biophysicists often rely on a simple, stochastic process and generate so-called *percolation clusters*.

Percolation clusters are generated by taking an empty (3D) matrix and randomly adding "obstacles" (non-zero) elements to the matrix. When the fraction of non-zero elements reaches a certain threshold, called the percolation threshold, the size distribution of the contiguously accessible elements (cluster) follows a power law and diffusion in such a cluster is anomalous (Ben-Avraham and Havlin 2000; Havlin and Ben-Avraham 1987). For each simulation setting, a new percolation cluster should be drawn.

Unfortunately, we did not have time to implement diffusion inside fractals in `simSPT`, although the simulation scheme is relatively easy to implement on a lattice.

6. **Fractional Brownian motion.** Fractional Brownian motion is the type of dynamics that arises when a particle diffuses and "bounces back" on a viscoelastic medium. In such a medium, anomalous diffusion arises from the interaction between the viscoelastic medium and the particle.

   In practice, fractional Brownian motion can equivalently be described as a generalization of the Brownian motion described above. In a traditional Brownian motion B, the increment are independent, which implies that the covariance of the process between two time points (s and t) is zero (assuming $s \neq t$):

   $$\mathbb{E}(B(s)B(t)) = \delta(s-t)$$

   In the case of a fractional Brownian motion (characterized by its Hurst parameter $H \in (0, 1]$), the increments are not independent, and the covariance between two time points is non-zero:

   $$\mathbb{E}(B_H(s)B_H(t)) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$$

   The Hurst parameter is directly related to the anomalous diffusion coefficient α, with $\alpha = 2H$. When $H > 0.5$, the process displays a positive correlation: if it tended to go in one direction at time t, it is more likely than by chance to go in the same direction at time $t + dt$. Conversely, the process displays anti-correlation if $H < 0.5$, and it it more likely to go in opposite direction at $t + dt$ than at time t. When $H = 0.5$, one exactly recovers Brownian motion.

   The simulation of fractional Brownian motion is not as straightforward as with Brownian motion or CTRW, because the definition of the fBm is implicit (the fBm is defined by its correlation properties). Indeed, many techniques have been developed to simulate fBm (Coeurjolly 2007; Shevchenko 2015) and a very exhaustive review in (Dieker 2004). The core difficulty lies in the fact that because the covariance is non-zero, one can say that this process has memory, which means that a fBm jump at time n can be computed given the position of the process at all times $[0, \ldots, n-1]$. Furthermore, the simulation usually requires to compute (and invert) the full covariance matrix

   They can broadly be divided in two classes:

   - Approximate simulation methods: they can either simulate a specific sub-type of fBm (not any fBm sequence of numbers can be produced), or rely on assumptions on how to compute/invert the covariance matrix. Most of them were developed for the case of $H > 1/2$, which is not applicable to motion in a viscoelastic medium and are usually inaccurate for short delta-t. These include the approximate circulant method, the Paxson method and wavelet-based methods (all reviewed in Dieker 2004).
   - Exact simulation methods: one can prove that these simulations exactly simulate fBm. They differ in the way the covariance matrix is computed/inverted, depending on the mathematical decomposition/reformulation that is being used. Since these are exact methods, they all produce the same fBm in the end, and only differ in the time/memory they require. Depending on the size of the fBm to be generated, one method or another could be more

efficient. Exact methods include the Hosking method, the Cholesky method and the Davies and Harte method. This latter relies on a Fast Fourier Transform to compute the covariance matrix, highly speeding up the simulation when long trajectories are to be simulated.

`simSPT` implements the fBm simulation library developed by Ton Diecker (Dieker 2004). We chose to use the Hosking method because it appeared faster for the size of problems we simulated (trajectories with less than 512 steps).

State changes are not implemented in `simSPT` when anomalous diffusion has been selected. Fractional Brownian motion has only one parameter, the Hurst parameter H (that is equal to half of the anomalous diffusion parameter, that is $H = \alpha/2$.

A small example is specified below:

```
./simSPT -motion=fbm -H=0.2 -D1=3.0 -p1=1.0 -n_traj=50000 -sigma=0
 -file=./outL2.txt -seed=22
```

1.1.4 Perspectives and conclusion

In this chapter, we presented a SPT simulation tool, `simSPT`, that balances ease-of-use with the incorporation of relevant biases usually encountered in nuclear SPT. In particular, the tool can reproduce the distribution of trajectory lengths encountered in regular SPT experiments. It can simulate one or several sub-populations of molecules undergoing Brownian motion, and can also incorporate interconversion between an arbitrary number of states. The tool was used to generate the simulations to benchmark Spot-On and other tools, as described in (Anders S. Hansen, Woringer, et al. 2018).

The strengths of `simSPT` are the fact that the tool is properly documented and packaged. Most of the options are available through the command-line, allowing to build complex simulations in one line of code, and to generate scripts to produce series of simulations. Second, because the tool is focused on SPT simulation, it requires a comparatively low amount of configuration, which make it easy to use. Third, as far as we know, `simSPT` is one of the fastest simulation tool available on the market. This was instrumental in the case of Spot-On, where we could benchmark our analysis tool against a wide range of conditions (inter-conversion, diffusion coefficients, localization error, axial detection range, etc). The tool is used by several members in the lab, and also by collaborators.

Moreover, the implementation of anomalous diffusion simulation routines make it instrumental for the complex diffusion challenge (section 4).

Finally, `simSPT` suffers some limitations. The first one is the limitation in terms of geometries that can be simulated: so far, only spherical and cubic geometries can be simulated. This limitation is easy to waive, since the routines confining the trajectories are independent of the type of motion (Szilvási-Nagy 1984). Another, slightly more intricate, is the fact that `simSPT` simulates out-of-equilibrium dynamics: particles are drawn at random from the several states, under the assumption that the pool of fluorescent proteins is never depleted. This is a reasonable approximation for short simulations, but is not realistic for longer ones. Furthermore, this approach yields out-of-equilibrium dynamics when clusters enriched in proteins are simulated, with a net flow from the cluster towards the exterior. These issues will have to be addressed in a new version of `simSPT`.

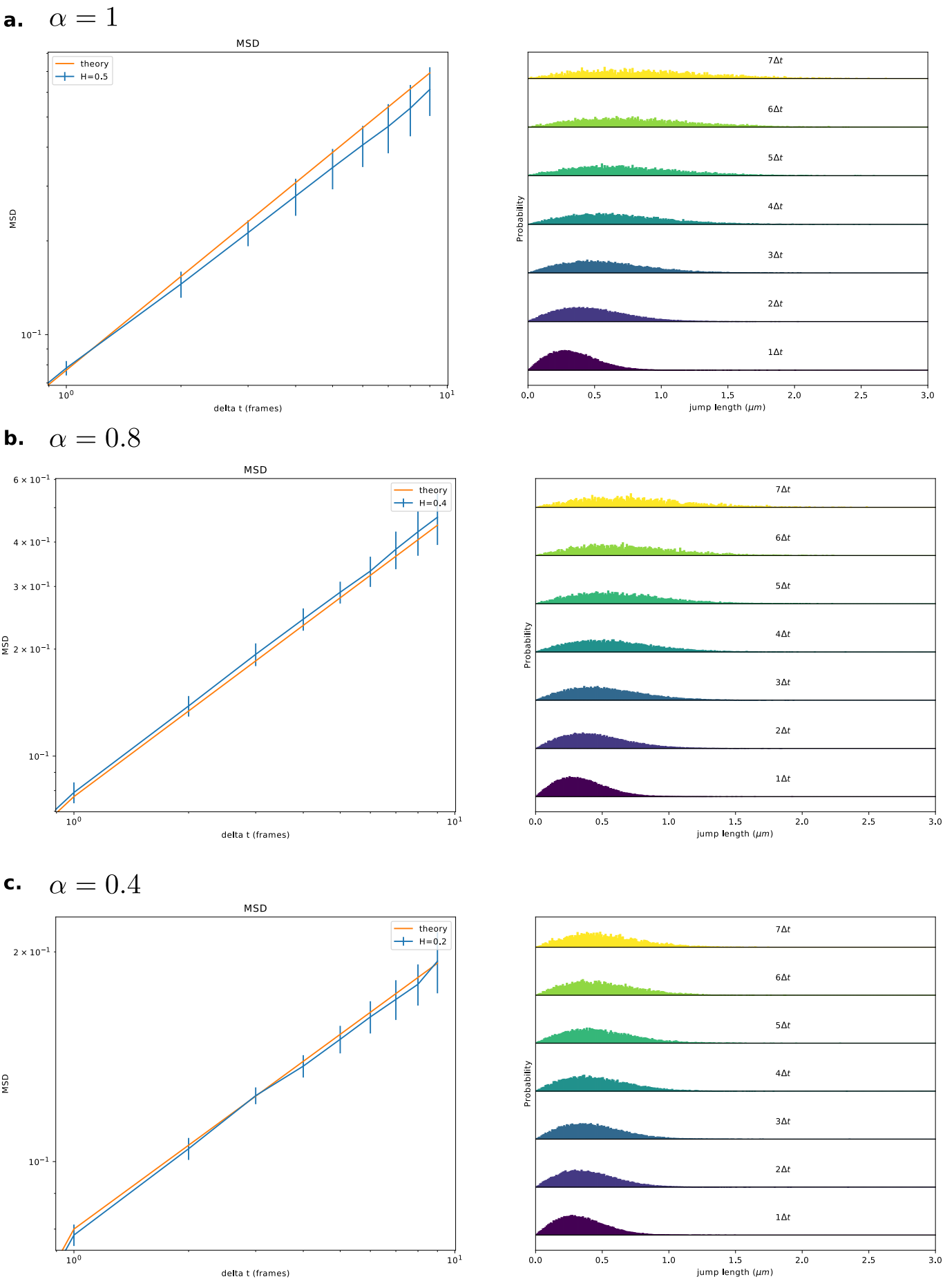


Figure 1.5: **Simulation of fractional Brownian motion**. (*left*) MSD computed from simulated data (blue) and corresponding theoretical MSD in log-log scale. (*right*) plot of the jump length distribution at various time intervals. **(a,b,c)** display three levels of anomalous diffusion (α=1.0, 0.8 and 0.4, respectively).

1.2 Better detection tools

Introduction. In the previous chapter (section 4.2) we introduced 2D SPT of 3D-diffusing proteins. In 2D SPT, the 3D motion of the observed protein is projected in 3D, and the motion in the x and y dimensions is recorded while the motion in the z dimension is lost (Figure 1.6a).

In practice, doing 2D SPT is relatively easy because most detection algorithms used only detect molecules that are "in focus" (within the focal plane of the objective) (for instance Sergé et al. 2008; Tinevez et al. 2017). Indeed, "in focus" particles appear as diffraction-limited spots of minimal width and maximum intensity. Conversely, as a particle moves out-of-focus, its point-spread function appears dimmer and more extended. Such molecule is usually excluded by the basic spot detectors used in 2D SPT (Figure 1.6b; in practice, it depends on the magnitude of the displacement and whether the detector assumes a PSF model). In these conditions, the observation window is usually limited in the third dimension to approximately 0.7-1.0 µm.

When the observation window is limited to 0.7-1.0 µm in depth, a diffusing protein moves out-of-focus very quickly, in the span of a few frames, depending on the diffusion coefficient (Figure 1.6a). All-in-all, regular 2D SPT data of proteins diffusing at diffusion coefficients around 1-3 µm²/s usually yield a mean trajectory length of ~ 4 frames, as can be seen in (Figure 1.6c).

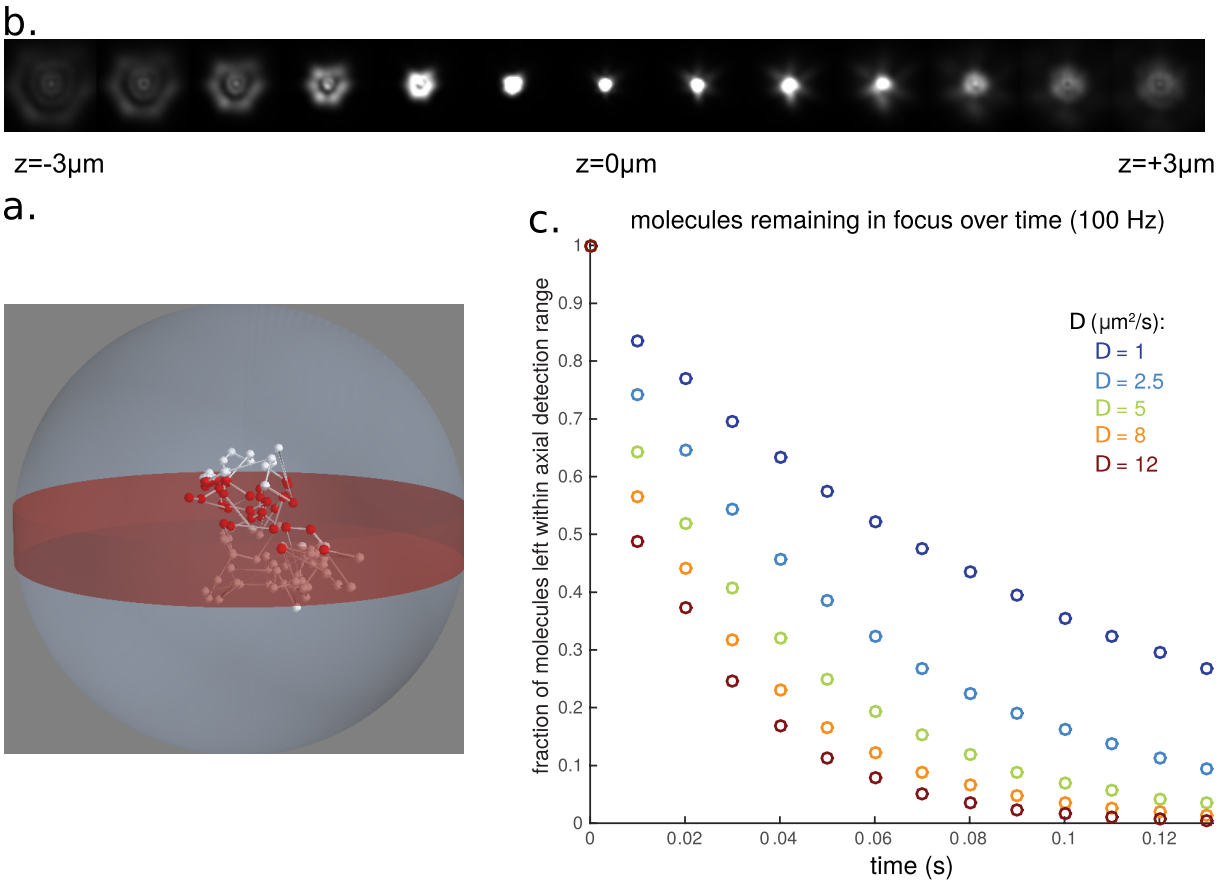


Figure 1.6: **Limitations of 2D SPT. (a).** Particles are observed in 2D while they diffuse in 3D, and they can move out-of-focus. **(b).** Experimental Airy PSF over 6 µm: only in-focus particles appear as bright spots (image courtesy Benoît Lelandais). **(c).** Plot of the number of molecules remaining in-focus as a function of time, plotted for various diffusion coefficient and assuming imaging at 100 Hz. Plot based on simulations of SPT.

With such short trajectories, it is extremely difficult to derive accurate statistics because of the high magnitude of the statistical noise (theoretically explored by Michalet and Berglund 2012. There is thus a need to extend the detection range in the third dimension. No matter the technique used, one

faces a fundamental trade-off when trying to extend the detection range in z: as the particle moves out-of-focus, the photons are progressively spread, and the PSF is progressively lost in the noise. To counter-balance this effect, the only solution seems to increase the exposure time. This is done at the expense of the PSF shape, because diffusing particle will have time to diffuse during one frame, as described in (section 4.2).

This section presents preliminary simulations and techniques in order to move from 2D to 3D fast SPT. This project was realized in very close collaboration with Benoît Lelandais from the Zimmer lab. Benoît developed an extension to ZOLA (Aristov et al. 2018) in order to simulate an arbitrary PSF at a given (x, y, z) position and he implemented several fitting routines that we had compared together. I generated the simulations.

We aim at solving the problem of detecting and localizing in 3D a particle that undergoes 3D diffusion under some level of motion blur. After a short introduction to motion blur (section 1.2.1) and 3D localization methods (section 1.2.2), we formulate the problem as a multi-emitter fitting problem (section 1.2.3) and propose both a simulation framework (section 1.2.4) and algorithms (section 1.2.5) to solve it. Preliminary results are also briefly presented.

1.2.1 Origins of motion blur

In cells, most of the transcription factors that have been studied display at least one sub-population that moves with a diffusion coefficient in the range 1-5 µm²/s (section 3.3). As detailed in (section 4.2) and assuming Brownian motion, the propagator equation gives the probability that a diffusing particle makes a jump of a given distance r in a time interval t_{exp} (equation 4.2). From this equation, one can easily compute the fraction of molecules that will move more than some number, r_{max}, during an exposure time, t_{exp}, given a free diffusion constant D using the following equation:

$$\mathbb{P}\left(r > r_{max}\right) = e^{-\frac{r_{max}^2}{4D_{free}t_{exp}}}$$

For example, if we define motion-blurring as moving more than 2 pixels ($>$ 320 nm assuming a 160 nm pixel size) during the excitation, an exposure time of 10 ms and a typical free diffusion constant of 3.5 µm²/s (e.g. Sox2), we get:

$$\mathbb{P}\left(r > r_{max}\right) = e^{-\frac{(0.32\mu m)^2}{4\times 3.5\mu m^2 s^{-1}\times 0.010 s}} \simeq 0.481$$

Thus, even for a relatively slowly diffusing protein, with a 10 ms exposure we should expect almost half (48%) of all free molecules to show significant motion-blurring.

Motion blur is usually avoided by using short excitation times, sometimes shorter than the camera acquisition time, a technique called stroboscopic illumination (J. Elf, G.-W. Li, and X. S. Xie 2007), as detailed in the spaSPT technique explained in (section 1.3) and in (Anders S. Hansen, Woringer, et al. 2018). Stroboscopic illumination and short exposure times both require that the laser pulses are concentrated within a very short time window (usually 0.5-5 ms), and thus necessitate powerful lasers. In several SPT acquisition settings (e.g: Teves, An, Anders S Hansen, et al. 2016; Anders S. Hansen, Pustova, et al. 2017; Anders S. Hansen, Woringer, et al. 2018; Boehning, Dugast-Darzacq, Rankovic, Anders S. Hansen, T.-K. Yu, et al. 2018), laser power is the factor limiting the decrease in exposure time.

Thus, a diffusing protein moves during the camera exposure. Following equation 4.2, one might think that this motion blur yields a Gaussian spread, that increases with the exposure time, since the equation describes a Gaussian with a standard deviation $\sigma = \sqrt{2Dt_{exp}}$ (Figure 1.7a). However, this is only true when averaging many trajectories, and it has been theoretically and experimentally shown that individual samples of a Brownian diffusion are never isotropic: a random walk statistically adopts a given form factor (its its ratio of dimensions, or tensor) or "shape" (Rudnick and Gaspari 1987; Sciutto 1994; Haber, Ruiz, and Wirtz 2000). This is exemplified in Figure 1.7b.

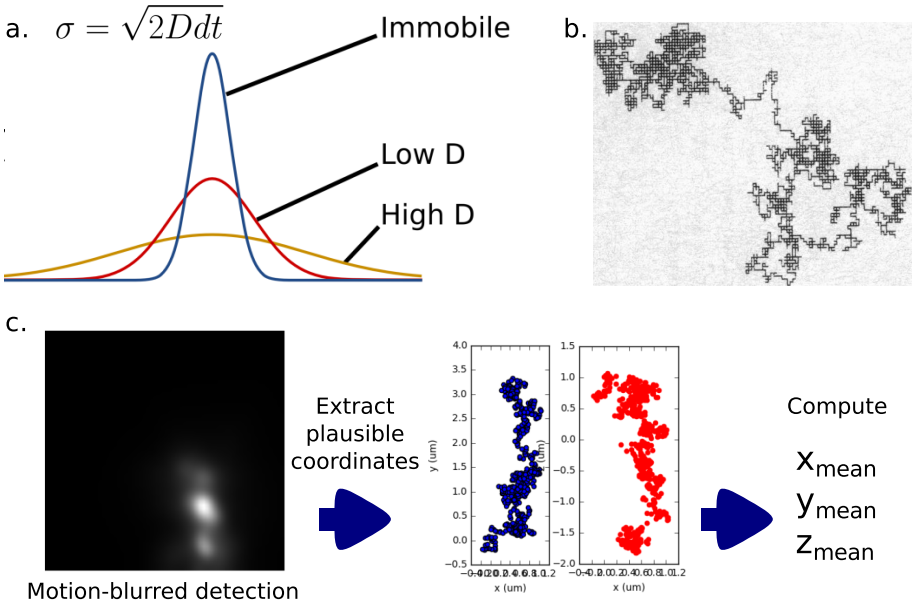


Figure 1.7: **Towards 3D SPT. (a)** In average, a random walk is symmetric and normally distributed, but **(b)** every single realization is actually asymmetric (figure from Rudnick and Gaspari 1987). **(c)** procedure to recover a 3D position from a 2D, motion-blurred detection. From a motion-blurred detection (*left*), the positions of the emitter are estimated (*center*) and the mean position in 3D is computed (*right*).

To sum up, when doing fast, 3D SPT, a certain amount of motion blur is unavoidable for the following reasons:

- The photons are more spread out than in 2D.
- Whereas in 2D, the exact shape of the PSF does not matter and precise 2D position can be recovered using simple centroid methods. (Deschout, Neyts, and Braeckmans 2012), in 3D the position in the third dimension is encoded in the shape of the PSF and small distortions (smaller than one pixel) might totally compromise the resolution in z.

In the following sections, we propose a framework to account for motion blur in 3D localization.

1.2.2 Introduction to 3D localization using PSF shaping

First, we briefly detail how 3D localization microscopy is usually performed. When a point source is observed through an objective, its image is not a point, but a diffraction spot called the point-spread function (PSF). Fourier optics can be used to mathematically determine the shape of the PSF, given the optical components of the microscope (Goodman 2005).

On an ideal microscope, the PSF is an Airy function, that is reasonably approximated by a symmetric Gaussian function. In practice, small deviations (aberrations) from the Airy PSF are present (Theer, Mongis, and Knop 2014) and one generally tries to avoid them. When PSF shaping is implemented, aberrations are introduced on purpose in the optical path, in order to make the PSF strongly asymmetric in z, allowing to encode the z position in the shape of the PSF. Initially this was implemented as an astigmatic PSF, in which the PSF appears elongated in one dimension for low z and the orthogonal direction for high z. Astigmatism can be performed using a cylindrical lens (B. Huang, W. Wang, et al. 2008). Later on, new engineered PSFs were developed using deformable mirrors (Izeddin, Beheiry, et al. 2012; Aristov et al. 2018), spatial light modulators and phase masks.

In parallel to the development of PSF shaping, many advances were performed in order to precisely localize individual fluorophores in 2D and 3D and take into account additional characteristics of the

microscope, and more complicated PSF than approximately Gaussians. This includes theoretical work to accurately determine the localization error (Mortensen et al. 2010; Rieger and Stallinga 2014) and the effect of fluorophore polarization (Enderlein, Toprak, and Selvin 2006), but also the design of entire localization algorithms (Basset et al. 2015; Aristov et al. 2018; Siemons et al. 2018; Yi, Piestun, and S. Weiss 2019; Fazel et al. 2019, compared in Sage, Pham, et al. 2018). The combination of both PSF shaping and efficient reconstruction algorithms allowed to reach a precision of $\sim$ 10 nm in x,y and $\sim$ 50 nm in z.

In the next section, we present how motion blur can be incorporated in order to recover 3D localizations.

1.2.3 Mathematical framework

1. **Problem formulation.** When an image of a freely-diffusing protein is made through a microscope with active PSF shaping, the following mathematical description can be made:

   (a) The particle diffuses with a diffusion coefficient D, and its 3D position X_t across time t can then be described by the following stochastic differential equation: $dX_t = \sqrt{2D}dB_t$, with B_t a standard Brownian motion.

   (b) The particle is observed during the exposure time of the camera, t_{exp}.

   (c) At a given time t, the particle emits photons and behaves as a point source. The image of this point source is then given by the (3D, shaped) PSF $\sigma_{PSF}(x, y, z)$ of the microscope.

   (d) When the photons arrive on the camera detector, the recorded "motion-blurred PSF" is then an image I of the motion blur due to the diffusion of the particle during t_{exp}, convolved by the 3D PSFs.

2. **Motion blur as a stochastic deconvolution problem.** Mathematically, the recorded image I at pixels (i, j) can then be described as, assuming constant intensity (photon emission) of the particle during T_{exp}:

$$I = \int_0^{T_{exp}} \sigma_{PSF}\left(X_t(x, y, z)\right) dt$$

   This can be recasted as a convolution problem, by defining the "image trajectory" $S_t \in \mathbb{R}^3$ of the random walk, namely the curve in 3D described by the random walk until time t:

$$S_t = \int_0^t \int_{\mathbb{R}^3} \mathbb{1}_{X_t} dx dy dz d\hat{t}$$

   Then, the resulting 2D image I can be expressed as a convolution with a 3D PSF:

$$I = S_t * \sigma_{PSF}$$

   The deconvolution problem aims at recovering S_t given σ_{PSF} and I. Three major difficulties arise when comparing this problem with a traditional deconvolution problem:

   - S_t should be 3D, whereas I is 2D only (the third dimension is encoded in the shape of the PSF.
   - S_t should be super-resolved compared to I, in order to achieve sub-pixelic localization.
   - S_t is the "image trajectory" of a random walk, and as such displays no regularity that could be exploited, as usually used in 3D deconvoluion algorithms.

3. **Motion blur as a multi-emitter fitting problem.** Thus, we realized that most of the deconvolution algorithms could not be used to perform such task. We also realized that our approach did not require the identification of the entire S_t at high resolution, but is sufficient to calculate some statistical descriptors of S_t. Indeed, our aim is for each motion-blurred PSF to recover only the mean location of the particle in (x, y, z), which is much less information than contained in S_t, which itself contains much less information than X_t.

   Indeed, the quantities that we are willing to extract are the mean position of the particle during the camera exposure time t_{exp}. We denote X_t^x (respectively X_t^y, X_t^z) the x (respectively y, z) component of the Brownian motion, and express the mean x position: $\bar{x}$:

$$\bar{x} = \frac{1}{t_{exp}} \int_0^{t_{exp}} X_{\hat{t}}^x d\hat{t}, \text{ respectively } \bar{y} \text{ and } \bar{z}$$

   We thus decided to reformulate the problem as a multi-emitter fitting problem. In such a setting, we assume that the observed image I can be reasonably expressed as the superposition of N (discrete) 3D PSFs. This is equivalent to discretizing the random walk. In other words, S_t is approximated by a discrete series of N points. Each of these points represents the average location of the initial walk X_t during a time t_{exp}/N. We denote the discretized walk as $\hat{X}_n$, with $n \in [0, N]$, and $\hat{X}_n$ defined as follows:

$$\hat{X}_n = \frac{N}{t_{exp}} \int_{n/N*t_{exp}}^{(n+1)/N*t_{exp}} X_t dt$$

   Formally, when N increases, then $\hat{X}_n$ should converge to X_t. In this setting, an estimate of $(\bar{x}, \bar{y}, \bar{z})$ is given by the average position of the N 3D points:

$$\bar{x} \simeq \frac{1}{N} \sum_{n=0}^{N} \hat{X}_n^x, \text{ respectively } \bar{y} \text{ and } \bar{z}$$

   This formulation provides two highlights:

   - First, it hints at a simulation methods: one can simulate Brownian motion with N substeps per exposure time, convolve them in 3D using a simulated 3D, shaped PSF and sum the resulting images in order to produce the image I.
   - Second, it hints at an inference technique, since the problem is now reduced to a parametric estimation of $3N$ parameters (x, y, z for each of the N points).

1.2.4 Data simulation

Once the problem was formulated as in (section 3), we decided to first adopt a simulation approach in order to:

1. Gain a better understanding of the effect of motion blur, and determine how symmetric are the PSF generated. Indeed, the trajectory of a random walks is known to be highly non-symmetric (Rudnick and Gaspari 1987; Sciutto 1994) and to adopt in 3D an aspect ratio in average $R_x, R_y, R_z = 11.80 : 2.69 : 1.00$. Furthermore, one can assume that the z information is lost as soon as the particle moves by more than half a pixel (80 nm) in any direction. The fraction of molecules in which the z information is lost in motion blur can then be approximately estimated using (equation 4.2; $\mathbb{P}(r > r_{max}) = e^{-\frac{r_{max}^2}{4Dt_{exp}}}$). For a typical transcription factor diffusing at 3 $\mu m^2/s$ exposed for $t_{exp} = 10ms$, this represents a 95% loss of molecules. If the frame rate is brought down to 2 ms, this value is still very high, about 75%. Coupled with simulations,

this first back-of-the-envelope calculation motivates the development of a multi-emitter fitting algorithm.

2. Determine the effect of the diffusion coefficient and the choice of the shape of the PSF on the quality of the estimation.

3. Provide a ground truth so that we can benchmark several fitting approaches.

1. **Choice of PSFs.** Several approaches exist to simulate realistic 3D PSFs, using various decompositions or optical models (Y. Li et al. 2017; Douglass 2017; *Fast and Accurate Three-Dimensional Point Spread Function Computation for Fluorescence Microscopy - Jizhou Li's Homepage* 2017). We used ZOLA (Zernike Optimized Localization Approach, described in Aristov et al. 2018) that incorporates an analytical generative PSF model that can generate any PSFs provided that its phase is continuous. ZOLA can infer the Zernike decomposition of a PSF using a z-stack of a bead. As such, one can gain a fully generative model of a given empirical PSF. We decided to use an extension of ZOLA developed by Benoît, in which the user can specify lists of (x, y, z) coordinates and the corresponding PSFs are simulated.

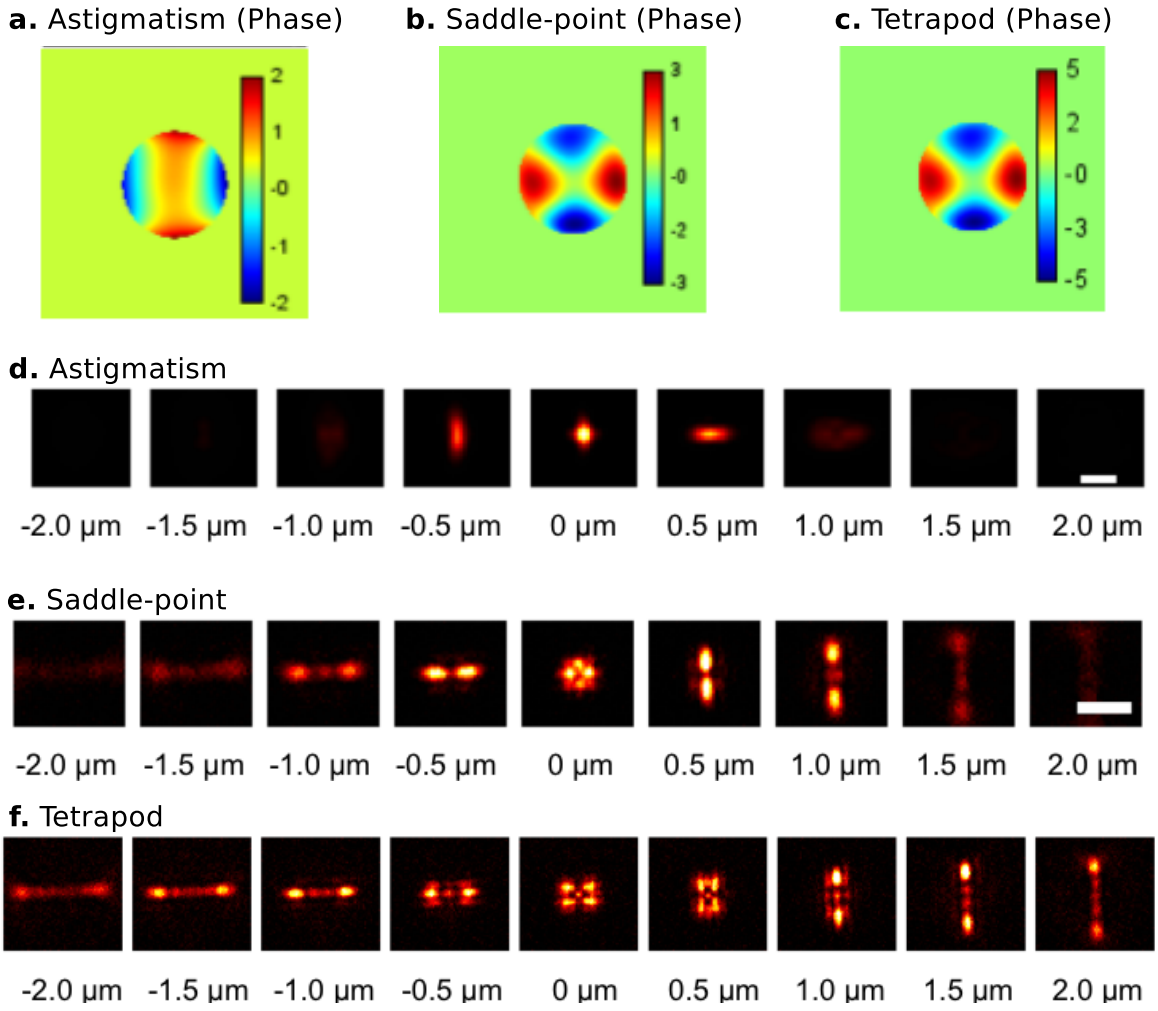


Figure 1.8: **Simulation of 3D, shaped PSFs. (a)** (respectively, **b**, **c**) Phase (at the back-focal plane of the microscope) corresponding to an astigmatic PSF (respectively, a saddle-point and a tetrapod PSF). **(d)** (respectively, **e**, **f**) Corresponding (experimental) z-stack of the astigmatic PSF, over a 4 µm range (respectively saddle-point PSF and tetrapod PSF). Source: Aristov et al. 2018.

We simulated four types of PSF: (1) a symmetrical, *Airy* PSF, (2) an astigmatic PSF, (3) a saddle-

point PSF and (4) a tetrapod PSF, pictured in Figure 1.8d-f. All these PSFs can be implemented in a microscope using a deformable mirror device (DMD), resulting in a shaped phase (Figure 1.8a-c). For the case of astigmatic PSF, a cylindrical lens can be used in replacement of the DMD.

2. **Trajectory simulation.** We simulated the motion of a protein during one exposure frame. We simulated a wide range of conditions and varied the following parameters (Table 1.1). For each condition, we simulated 500 substeps and performed 100 replicates. For instance, when simulating one exposure frame of 10 ms, Brownian diffusion was simulated on 500 substeps of 0.02 ms (10 ms/500).

Table 1.1: Parameters used for the simulations of motion-blurred PSFs.

Parameter	name	unit	range explored	number of conditions
Exposure time	dt	ms	1-100	14
Diffusion coefficient	D	$\mu m^2/s$	0.1-10	8

In total, we generated 11200 trajectories. Then, for each of the 500 simulated (x, y, z) positions per trajectory, we used ZOLA to produce a corresponding PSF, using the four PSF shapes mentioned above. Finally, for each trajectory, the 500 PSF images were summed in order to produce motion-blurred images (Figure 1.9).

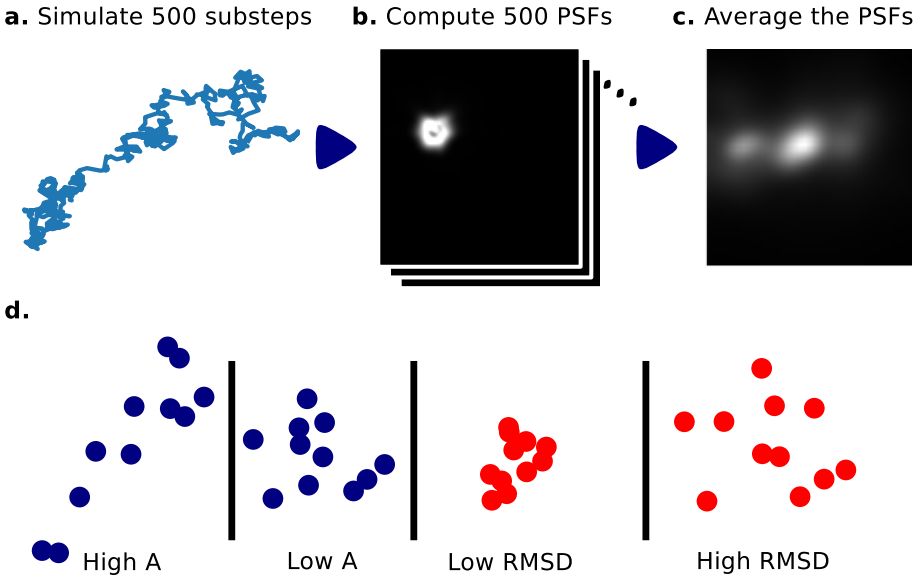


Figure 1.9: **Procedure to simulate and validate 3D SPT. (a)** 500 substeps of a trajectories are simulated, for a total time of one exposure frame, **(b)** 500 PSFs are computed using a realistic model used in ZOLA-3D, **(c)** The 500 frames are summed, and noise can be further added. **(d)** Illustration of two metrics used to characterize the trajectories: the asphericity (A) and the root mean square deviation (RMSD).

3. **Trajectory statistics.** We then characterized the simulated trajectories. Our goal was to get an overview of the simulated dataset, but also to select PSF that are expected to be "hard to fit". How to define an "Hard-to-fit" PSF? We decided to rely on two criterion applied to the underlying trajectory (illustrated in Figure 1.9d): the *root mean squared deviation* ($RMSD$) and the *asphericity* (A). The RMSD quantifies the spread of the detections whereas the asphericity quantifies whether the trajectory appears elongated in one dimension.

The RMSD is defined as: $RMSD = \sqrt{\frac{1}{N}\sum_{t=1}^{N}((\hat{x}-x_t)^2+(\hat{y}-y_t)^2+(\hat{z}-z_t)^2)}$.

The asphericity is defined as follows (Sciutto 1994; eq. 12): $A = \frac{1}{d-1} \frac{\sum_{i=1}^{d-1} \sum_{j=i+1}^{d} (\lambda_i - \lambda_j)^2}{\sum_{i=1}^{d} \lambda_i}$ with the λ_i the eigenvalues of the inertia matrix, and $d = 3$ for the 3D case.

The distribution of asphericities and radii of gyration is represented in Figure 1.10. The fact that most trajectories exhibit an asphericity higher than 0.5 suggests that motion blur should be accounted for in order to perform a fit of good quality, and that it cannot be neglected.

For the rest of the analysis, we mainly focused our efforts on the 10 trajectories with the highest asphericity and the 10 trajectories with the highest RMSD (Figure 1.10c). If the algorithm performs well on such "bad-behaved" examples, we expect that the algorithm will perform well generally (and further validations will be required).

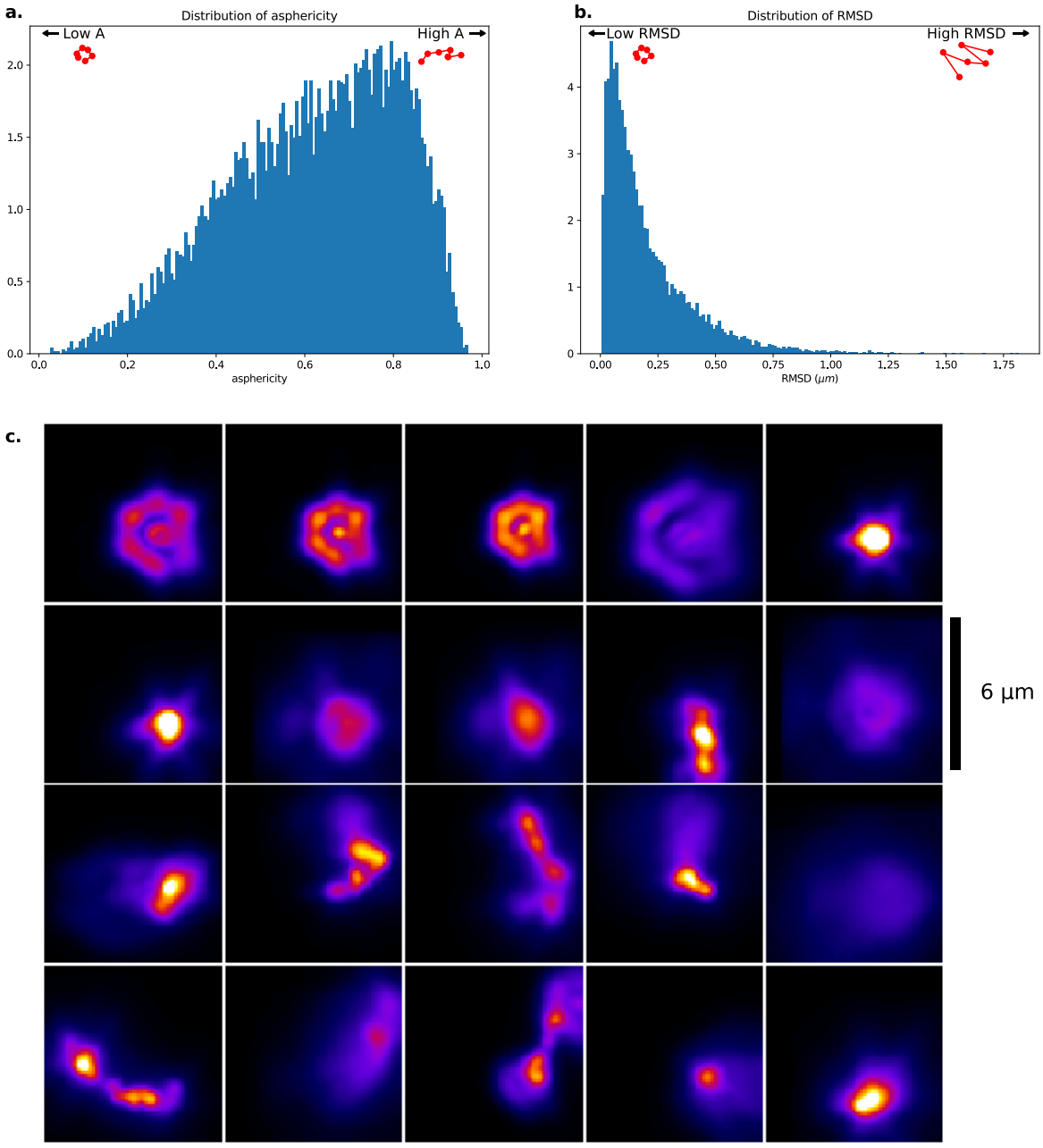


Figure 1.10: **Statistics of simulated trajectories and PSFs.** **(a)** Distribution of asphericity and **(b)** RMSD for the simulated trajectories. On average, the trajectories exhibit a relatively high asphericity (with a mode around 0.8) and a highly variable RMSD. **(c)** The 20 trajectories with the highest RMSD (two top rows) and the highest A (two bottom rows).

4. **Noise models.** As a first attempt, we assumed a noiseless model and fitted the data assuming an infinite signal-to-background ratio. Then, in order to simulate more realistic imaging conditions, Poissonian noise was added to the images. We tested one noise condition, consisting in $n_p = 1000$ photons per PSF and a background of $n_{bg} = 1$ photon per pixel.

   Thus, the noisy image I_{noise} is a realization of a Poisson process whose intensity derives from I. We now assume that I is normalized (the sum of all the pixel values is 1). Representative images are plotted in Figure 1.11.

   Then, I_{noise} can be expressed in the units of number of photons as follows:

$$I_{noise} = \mathcal{P}\left(n_p * I + n_{bg}\right)$$

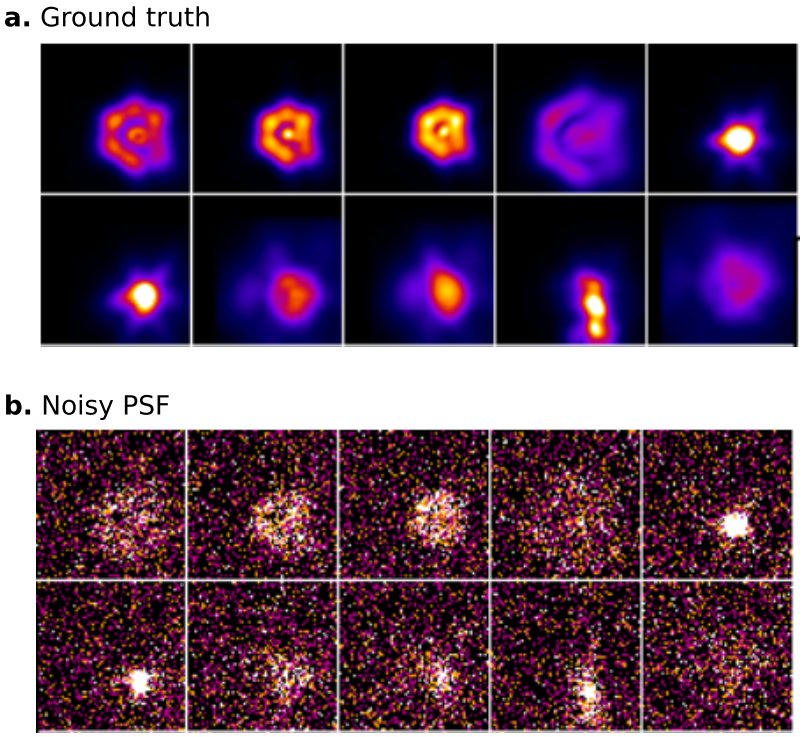


Figure 1.11: **Example of noisy PSFs. (a)** 10 simulated PSFs and **(b)** the same PSFs, with noise added (1000 photons and background of 1 photon/pixel).

1.2.5 Fitting

Once the data has been simulated, one can run a fitting algorithm in order to estimate the position of n PSFs and recover $3N$ parameters. In practice, before choosing an algorithm, one should be aware that depending on the the PSF, identifiability issues may arise: more than one solution can come out of the optimization routine (section 2). Second, multi-emitter fitting is known to be more challenging than single-emitter fitting due to both the dimensionality of the problem and the existence of many symmetries, that lead to non-unique solutions, and adequate optimization algorithms are needed. We present below some results both in the noiseless and noisy cases.

1. **Identifiability issues.** When performing 3D multi-emitter fitting, the choice of the PSF is critical. Indeed, it might be very often impossible to distinguish some localizations if the PSF is not identifiable. An example is provided in (Figure 1.12), in which an astigmatic PSF was used. In that case, it is impossible to decide whether one observes an in-focus particle undergoing directional motion in the vertical direction or an immobile, out-of-focus particle. More complex PSF shapes reduce (but do not abolish) the identifiability issue.

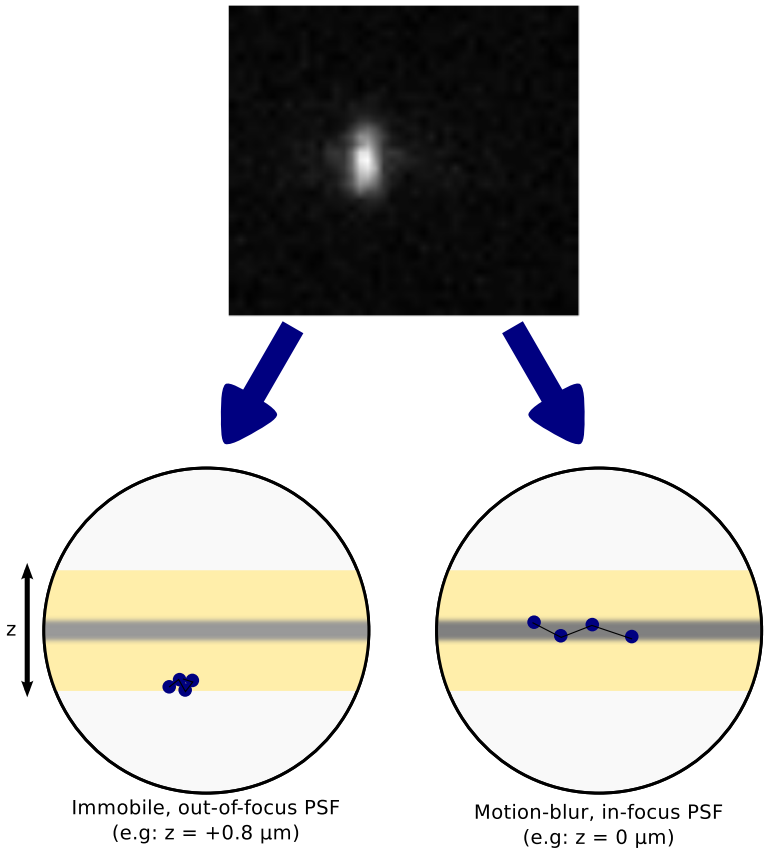


Figure 1.12: Identifiability of an astigmatic PSF: provided a spot such as as the one presented at the top of the figure, it is extremely difficult to state whether it corresponds to an out-of-focus, immobile particle (*bottom left*), or to a mobile, in-focus particle (*bottom right*).

2. **Choice of algorithm.** In order to solve the multi-emitter fitting problem presented above, one needs to choose an optimization algorithm. Using non-linear least-squares fitting is usually a good, by-default choice. However, in the case of multi-emitter fitting, we decided to opt for other techniques. The optimization problem to solve has many symmetries, and thus many local minima because the emitters fitted can be swapped. Thus, if $\mathcal{S} = (X_1, \ldots, X_N)$ is a solution to the optimization problem (where $X_i = (x_i, y_i, z_i)$ the coordinates of one of the emitters), then any permutation of $\mathcal{S}$ is also a solution of the problem. In other words, when the problem is solved with N emitters, it has $N!$ equivalent solutions. Such a high number of solutions is known to cause problems to non-linear least square algorithms.

   We instead decided to focus our work on algorithms known to perform well on non-convex problems, namely cross-entropy optimization and genetic algorithm (Botev et al. 2013; Benham et al. 2015; Lovinger 2018).

   Cross-entropy is an iterative stochastic optimization algorithm in which a population of candidate solutions are first sampled according to an initial distribution. Second, the quality of the fit of each of these candidate solutions is evaluated by computing the Kullback-Leibler divergence between the model corresponding to the candidate solution and the experimental data. Third, the 10% best-fitting candidate solutions are kept and their mean parameter value and variance is computed. These mean values and variance are then used to sample a new population of candidate solutions, and a new iteration can start. As the algorithm iterates, the variance progressively converges, and the mean value of the parameters of the models converge to accurate parameter estimates.

   Genetic algorithm is another iterative stochastic optimization in which a population of candidate

solutions is sampled, and the best fitting solutions are selected for the next round. They are allowed to perform two sets of moves: a local optimization, and recombination between good-fitting solutions. In practice, it is difficult to perform efficient recombinations because of the high number of symmetries of the problem. We partially tackled this problem by defining an arbitrary order to the emitters, using a nearest-neighbour search.

3. **Noiseless case.** On some preliminary simulations, the cross-entropy method provided better results, and the following results were obtained using this method. We first assessed the performance of our multi-emitter fitting algorithm on noiseless, motion-blurred PSFs. As described above, we used a subset of PSFs that displayed either a high RMSD or a high asphericity. Visually, the fit provided PSFs that resembled the ground truth (Figure 1.13).

   We then quantitatively assessed the quality of the fit. To do so, the mean position of the PSF was computed on both the ground truth and the reconstruction, and the absolute median error was computed as:

$$\text{median}_{(1..100)} \left| \frac{1}{5} \sum_{i=1}^{5} \hat{x_i} - \frac{1}{500} \sum_{j=1}^{500} x_i \right|$$

   The quantitative results are presented in Figure 1.14. For all the dimensions, the mode of the recovered median error on the position is around 5 nm, which is encouraging. A closer examination of the distribution of the error as a function of the diffusion coefficient and the exposure time reveals that for high diffusion coefficients and exposure times, the localization error greatly increases, reaching down to 350 nm for some cases.

   To provide a more objective comparison, we fitted the same motion-blurred PSFs using a single-emitter fitting routine: ZOLA-3D, and tried to see whether the multi-emitter fitting improved the estimates. The results are presented in Figure 1.15. A close comparison between the two techniques so far does not prove the superiority of the multi-emitter fitting algorithm. The single-emitter fit seems to always outperform the multi-emitter fit, even in cases of high diffusion coefficient and long exposure time.

   Several reasons might explain why the multi-emitter fitting is under-performing, including the long convergence of the cross-entropy algorithm compared to the optimization scheme implemented in ZOLA and the difficulty to initialize properly the position of the multiple emitters.

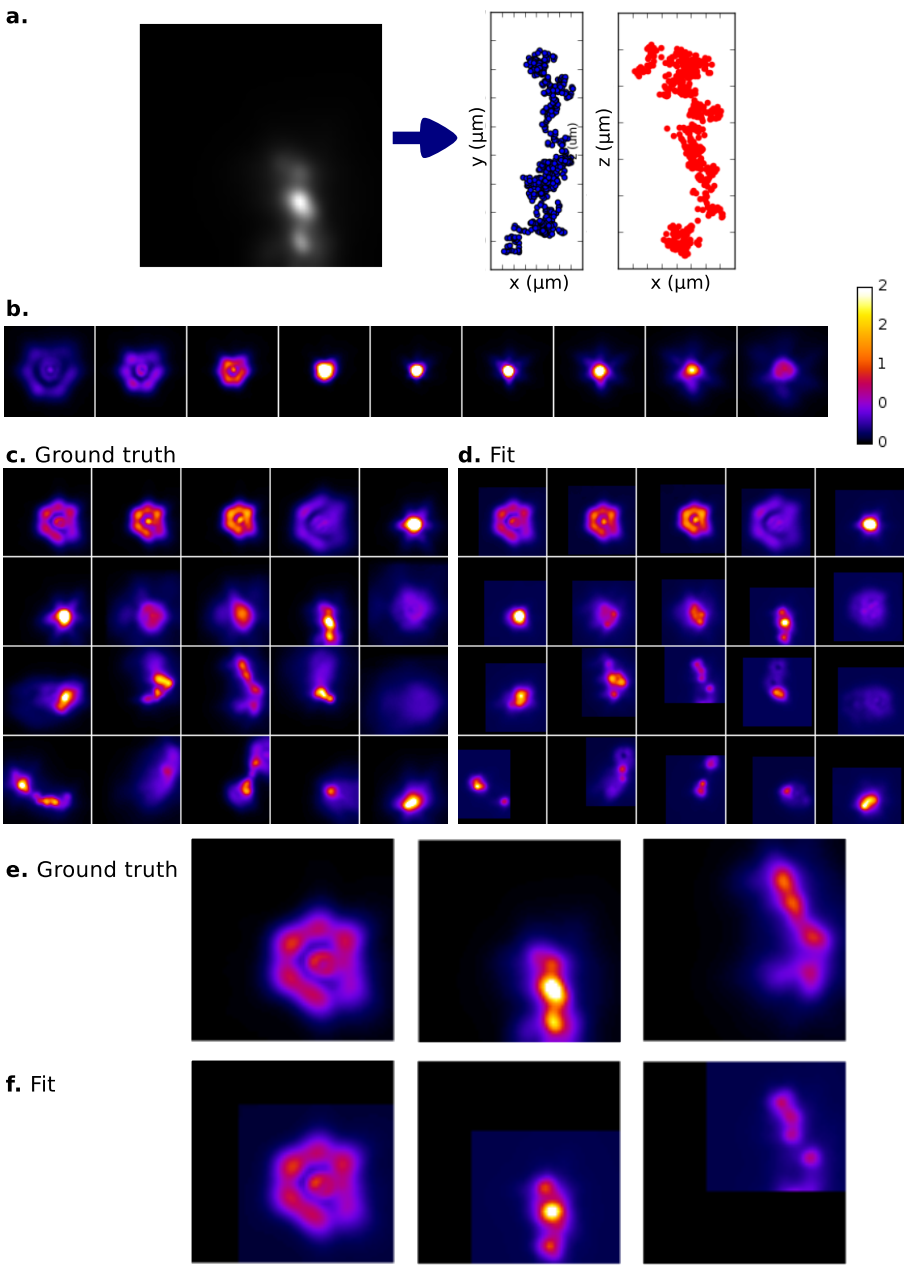


Figure 1.13: **Multi-emitter fitting in the noiseless case. (a)** Schematic of the problem solved: a cross-entropy algorithm is used to recover the (x, y, z) localization of single-point emitters from a motion-blurred PSF. **(b)** PSF used in this example: an Airy PSF. **(c)** Ground truth motion-blurred PSFs used as input. **(d)** Fitted multi-emitter PSFs provided by the optimization algorithm. **(e) and (f)** Zoom-in on some of the fitted PSFs. Fits were performed using five emitters.

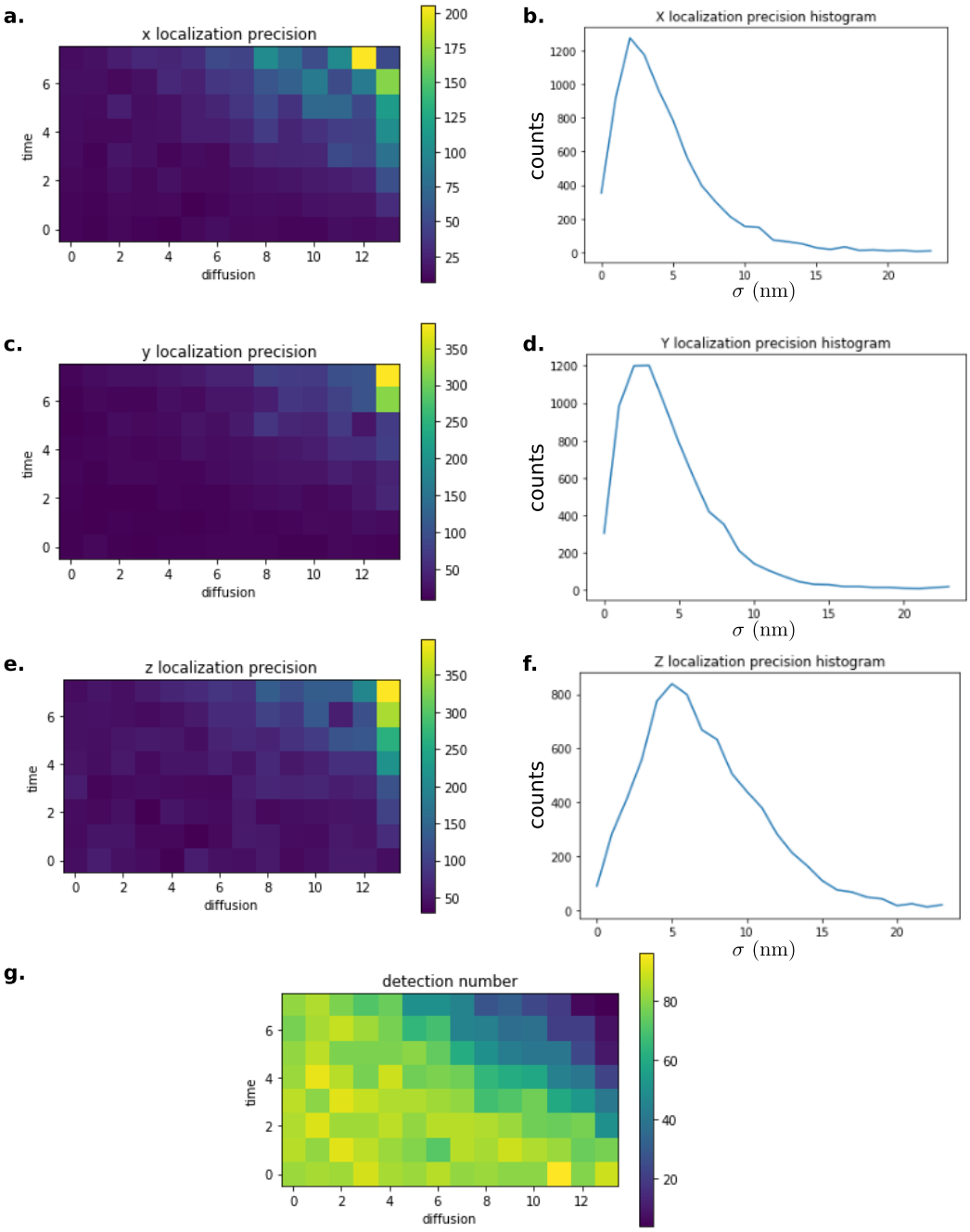


Figure 1.14: **Quantification of the accuracy of the multi-emitter fitting approach. (a,c,e)** median localization error in (x, y, z, respectively) as a function of the exposure time (*rows*) and the diffusion coefficient (*columns*, in $\mu m^2/s$). **(b,d,f)** Histogram of the localization precision in (x, y, z, respectively). For some motion-blurred PSF, not all emitters could be detected and **(g)** presents the number of detected motion-blurred PSFs. Fits were performed using five emitters.

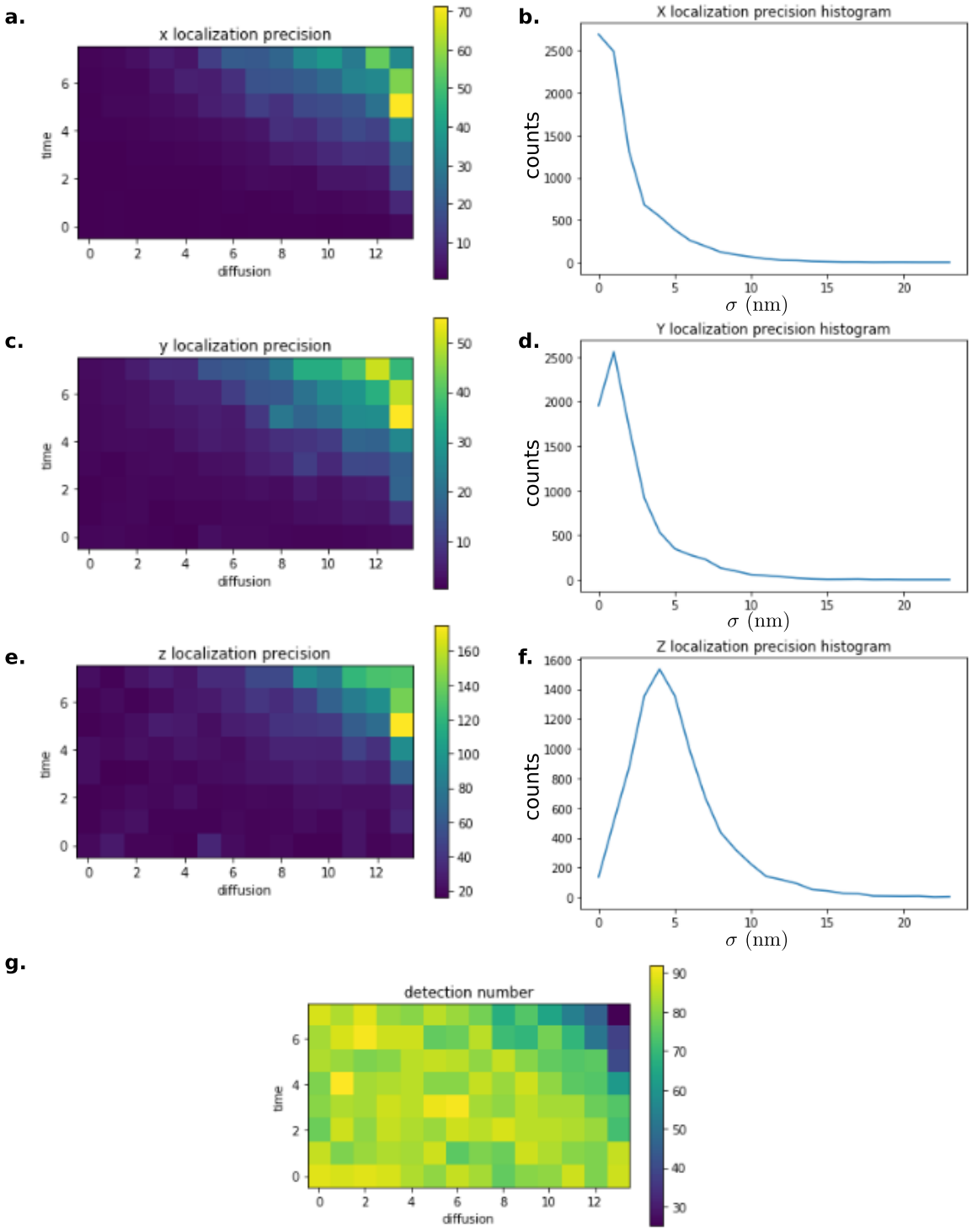


Figure 1.15: **Quantification of the accuracy of the single-emitter fitting approach. (a,c,e)** median localization error in (x, y, z, respectively) as a function of the exposure time (*rows*) and the diffusion coefficient (*columns*, in $\mu m^2/s$). **(b,d,f)** Histogram of the localization precision in (x, y, z, respectively). For some motion-blurred PSF, not all emitters could be detected and **(g)** presents the number of detected motion-blurred PSFs.

4. **Noisy case.** As an extra test, we also assessed the performance of the multi-emitter fitting algorithm in the presence of noise. The results are presented in Figure 1.16. Clearly, symmetric, out-of-focus PSFs tend to be reasonably fitted, however, highly motion-blurred PSFs appear poorly fitted, with the algorithm getting confused by the motion of the particle.

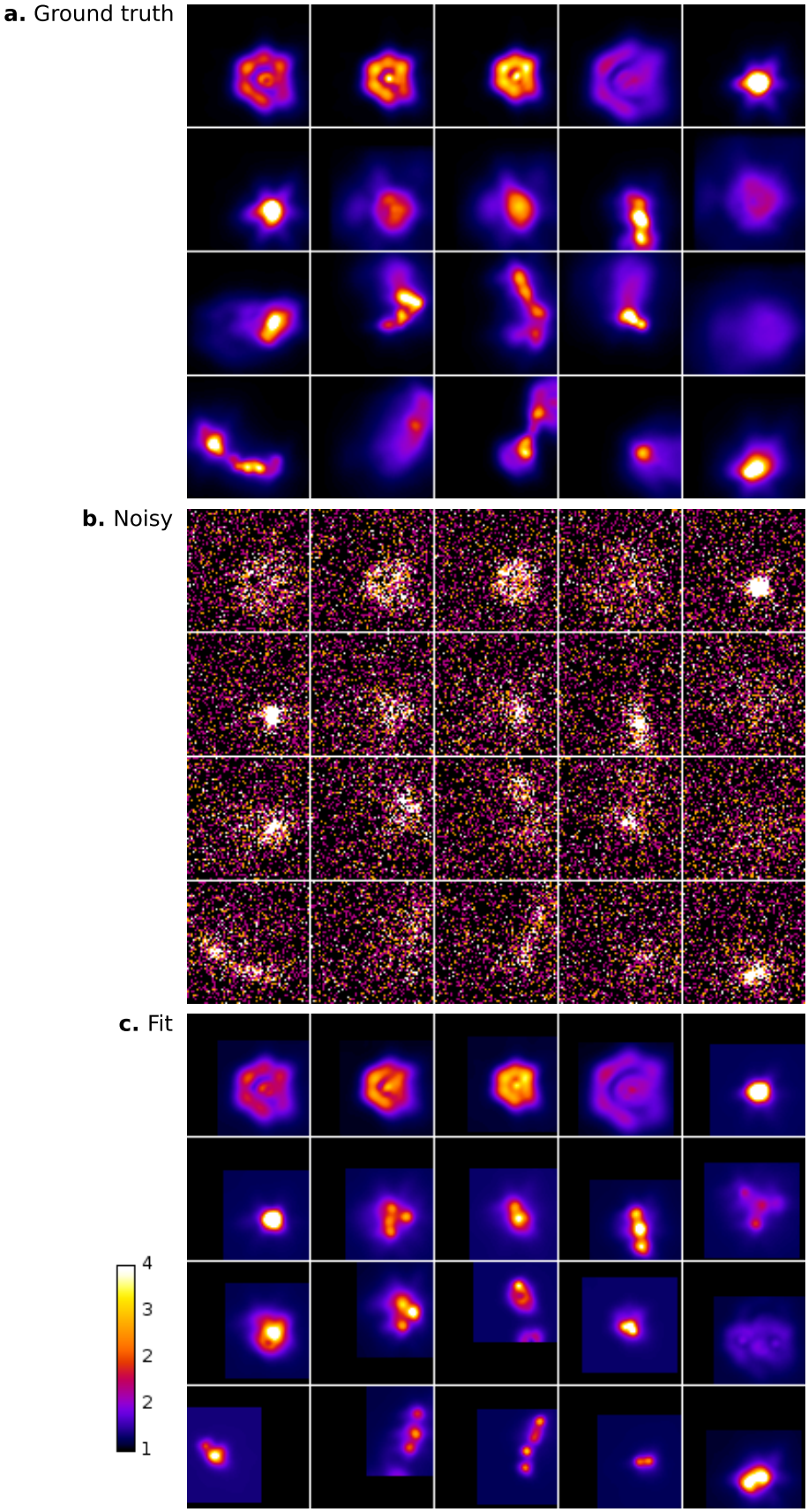


Figure 1.16: **Quality of the multi-emitter fitting in the presence of noise.** **(a)** Ground truth PSFs. **(b)** Noisy PSFs. **(c)** Fitted PSFs.

1.2.6 Conclusion

In this section we presented a multi-emitter fitting approach to recover the 3D spatial position of a diffusing particle under motion blur using PSF shaping. We provided a theoretical framework and a proof-of-concept implementation of an algorithm based on cross-entropy optimization. We benchmarked it against state-of-the-art PSF fitting algorithms in noiseless and noisy conditions.

Our approach can successfully recover the mean position of a motion-blurred particle, even in the case of high diffusion coefficients and long exposure times. However, when compared with ZOLA (Aristov et al. 2018), our method under-performs this single-emitter fitting technique. It is not clear why this is the case, but issues related to convergence and initialization of the cross-entropy routine might be at play and will have to be resolved.

Building on these multi-emitter fitting approach, and assuming a Brownian diffusion model, it is possible, at least theoretically, to infer an estimate of the diffusion coefficient of the particle. Indeed, metrics such as the mean maximal excursion method are typically designed to analyze the type of information recovered by the algorithm (Tejedor et al. 2010). Such a setting could also be coupled with fluorescence lifetime imaging (FLIM), in order to extract both the instantaneous single-molecule diffusion coefficient and information about the local environment, as detailed in (Bouchet et al. 2019).

To our knowledge, this problem has never been considered before. However, some related problems need to be mentioned. First, (Spille et al. 2012) performed 3D mRNA tracking in yeast using PSF shaping. In this imaging setting, motion blur is minimal because the studied mRNA diffuses slowly. Second, recent deep learning techniques were proposed to perform high density super-resolution detection (Nehme, L. E. Weiss, et al. 2018; Nehme, Hershko, et al. 2019). High-density super-resolution detection is identical to multi-emitter fitting, and as such, algorithms that perform well in one setting are expected to perform well in the other setting. To assess whether the approach introduced in (Nehme, Hershko, et al. 2019) can be used for live-cell imaging, we contacted the Shechtman lab to initiate a collaboration.

1.3 Inference tools

The detection and PSF-fitting tool presented in the previous section (section 1.2) is still very preliminary, and the acquisition of "long" trajectories remains an ethered dream. By long we mean "long enough so that single-trakectory statistics could be extracted" (Michalet and Berglund 2012). As a consequence, we decided to develop tools that would allow non-programmer users to extract more information from current state-of-the-art SPT data. To do so, we developed two approaches:

- Spot-On, a web interface and accompanying Python/Matlab codes that infer multi-population diffusion models based on the approach developed by (D. Mazza et al. 2012) and refined in (Anders S. Hansen, Pustova, et al. 2017). The published version of Spot-On is presented in this section and current developments in section IV.2.
- We developed a second approach that integrates information about the angle distribution of displacements in the propagator (section IV.3).

1.3.1 Summary of the work

Single-particle tracking (SPT) has become an important method to bridge biochemistry and cell biology since it allows direct observation of protein binding and diffusion dynamics in live cells. However, accurately inferring information from SPT studies is challenging due to biases in both data analysis and experimental design. To address analysis bias, we introduce 'Spot-On', an intuitive web-interface. Spot-On implements a kinetic modeling framework that accounts for known biases, including molecules moving out-of-focus, and robustly infers diffusion constants and subpopulations from pooled single-molecule trajectories. To minimize inherent experimental biases, we implement and validate stroboscopic photo-activation SPT (spaSPT), which minimizes motion-blur bias and tracking errors. We validate Spot-On using experimentally realistic simulations and show that Spot-On outperforms other methods. We then apply Spot-On to spaSPT data from live mammalian cells spanning a wide range of nuclear dynamics and demonstrate that Spot-On consistently and robustly infers subpopulation fractions and diffusion constants.

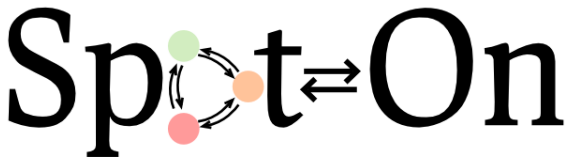


Figure 1.17: Logo of Spot-On

In this section, we present Spot-On. The development of a SPT analysis technique and web interface was done in close collaboration with Anders S. Hansen. He initially extended and reimplemented an approach previously developed. Together, we further extended the code and performed extensive validations, both on simulations and experimental data. Anders took care of the Matlab version, and generated 1064 single-particle tracking experimental datasets. I ported Spot-On to Python and to a web-interface. The experimental datasets are freely available in Spot-On readable Matlab and CSV file formats in the form of SPT trajectories at Zenodo. The experimental data is available at: `https://zenodo.org/record/834781`). I wrote `simSPT` (section 1.1), the code to produce the simulations used to validate Spot-On and explore its accuracy over a large range of parameters. The simulations are available in Matlab format at: `https://zenodo.org/record/835541`; The simulations are available in CSV format at: `https://zenodo.org/record/834787`. This section borrows entire paragraphs from (Anders S. Hansen, Woringer, et al. 2018) and from the Spot-On documentation.

1.3.2 Introduction.

Advances in imaging technologies, genetically encoded tags and fluorophore development have made single-particle tracking (SPT) an increasingly popular method for analyzing protein dynamics (Z. Liu, L. Lavis, and Eric Betzig 2015). Recent biological applications of SPT have revealed that transcription factors (TFs) bind mitotic chromosomes (Teves, An, Anders S Hansen, et al. 2016), how Polycomb interacts with chromatin (Zhen et al. 2016), that 'pioneer factor' TFs bind chromatin dynamically (Swinstead et al. 2016), that TF binding time correlates with transcriptional activity (Loffreda et al. 2017) and that different nuclear proteins adopt distinct target search mechanisms (Izeddin, Récamier, et al. 2014; Rhodes et al. 2017). Compared with indirect and bulk techniques such as Fluorescence Recovery After Photobleaching (FRAP) or Fluorescence Correlation Spectroscopy (FCS), SPT is often seen as less biased and less model-dependent (Goulian and Simon 2000; Mueller, Stasevich, et al. 2013; Shen et al. 2017). In particular, SPT makes it possible to directly follow single molecules over time in live cells and has provided clear evidence that proteins often exist in several subpopulations that can be characterized by their distinct diffusion coefficients (Mueller, Stasevich, et al. 2013; Shen et al. 2017). For example, nuclear proteins such as TFs and chromatin binding proteins typically show a quasi-immobile chromatin-bound fraction and a freely diffusing fraction inside the nucleus.

However, while SPT of slow-diffusing membrane proteins is an established technology (Weimann et al. 2013), 2D-SPT of proteins freely diffusing inside a 3D nucleus introduces several biases that must be corrected for in order to obtain accurate estimates of subpopulations.

- First, while a frame is acquired, fast-diffusing molecules move and spread out their emitted photons over multiple pixels causing a 'motion-blur' artifact (Berglund 2010; Deschout, Neyts, and Braeckmans 2012; Frost, H. E. Lu, and Blanpied 2012; Goulian and Simon 2000; Izeddin, Récamier, et al. 2014), whereas immobile or slow-diffusing molecules resemble point spread functions (PSFs; Figure 1.18A). This results in under-counting of the fast-diffusing subpopulation.

- Second, high particle densities tend to cause tracking errors when localized molecules are connected into trajectories. This can result in incorrect displacement estimates (Figure 1.18B).

- Third, since SPT generally employs 2D imaging of 3D motion, immobile or slow-diffusing molecules will generally remain in-focus until they photobleach and therefore exhibit long trajectories, whereas fast-diffusing molecules in 3D rapidly move out-of-focus, thus resulting in short trajectories (we refer to this as 'defocalization'; Figure 1.18C). This results in a time-dependent under-counting of fast-diffusing molecules (Goulian and Simon 2000; Kues and Kubitscheck 2002). Fourth, SPT analysis methods themselves may introduce biases; to avoid this, an accurate and validated method is needed (Figure 1.18D).

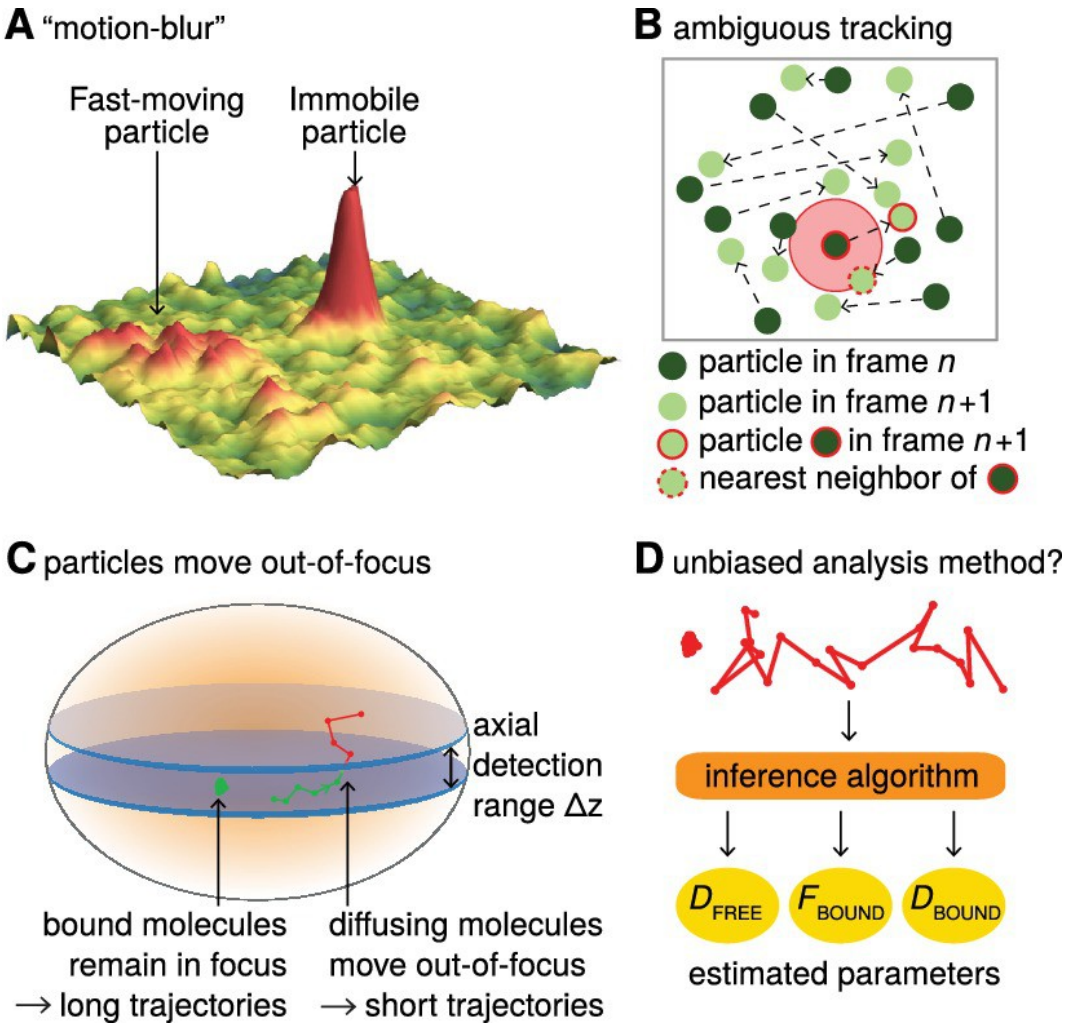


Figure 1.18: **Bias in single-particle tracking (SPT) experiments and analysis methods.** (A) 'Motion-blur' bias. Constant excitation during acquisition of a frame will cause a fast-moving particle to spread out its emission photons over many pixels and thus appear as a motion-blur, which make detection much less likely with common PSF-fitting algorithms. In contrast, a slow-moving or immobile particle will appear as a well-shaped PSF and thus readily be detected. (B) Tracking ambiguities. Tracking at high particle densities prevents unambiguous connection of particles between frames and tracking errors will cause displacements to be misidentified. (C) Defocalization bias. During 2D-SPT, fast-moving particles will rapidly move out-of-focus resulting in short trajectories, whereas immobile particles will remain in-focus until they photobleach and thus exhibit very long trajectories. This results in a bias toward slow-moving particles, which must be corrected for. (D) Analysis method. Any analysis method should ideally avoid introducing biases and accurately correct for known biases in the estimation of subpopulation parameters such as D_{FREE}, F_{BOUND}, D_{BOUND}. Source: `https://doi.org/10.7554/eLife.33125.003`.

Here, we introduce an integrated approach to overcome all four biases. The first two biases must be minimized at the data acquisition stage and we describe an experimental SPT method to do so (spaSPT), whereas the latter two can be overcome using a previously developed kinetic modeling framework (Anders S. Hansen, Pustova, et al. 2017; D. Mazza et al. 2012) now extended and implemented in Spot-On. Spot-On is available as a web-interface (`https://SpotOn.berkeley.edu`) as well as Python and Matlab packages.

1.3.3 Overview of Spot-On

Spot-On is a user-friendly web-interface that pedagogically guides the user through a series of quality-checks of uploaded datasets consisting of pooled single-molecule trajectories. It then performs kinetic model-based analysis that leverages the histogram of molecular displacements over time to infer the fraction and diffusion constant of each subpopulation (Figure 1.19). Spot-On does not directly analyze raw microscopy images, since a large number of localization and tracking algorithms exist that convert microscopy images into single-molecule trajectories (for a comparison of particle tracking methods, see (Chenouard et al. 2014); moreover, Spot-On can be one-click interfaced with TrackMate (Tinevez et al. 2017), which allows inspection of trajectories before uploading to Spot-On).

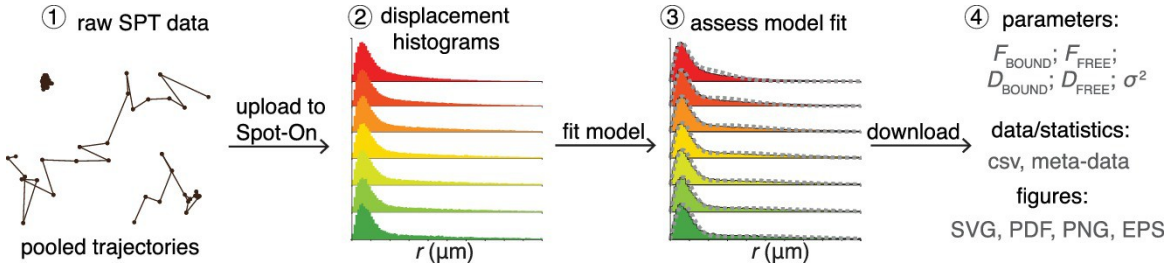


Figure 1.19: **Overview of Spot-On interface.** To use Spot-On, a user uploads raw SPT data in the form of pooled SPT trajectories to the Spot-On web-interface. Spot-On then calculates displacement histograms. The user inputs relevant experimental descriptors and chooses a model to fit. After model-fitting, the user can then download model-inferred parameters, meta-data and download publication-quality figures. Source: `https://doi.org/10.7554/eLife.33125.004`

To use Spot-On, a user uploads their SPT trajectory data in one of several formats (Figure 1.19). Spot-On then generates useful meta-data for assessing the quality of the experiment (e.g. localization density, number of trajectories etc.). Spot-On also allows a user to upload multiple datasets (e.g. different replicates) and merge them. Spot-On then calculates and displays histograms of displacements over multiple time delays. The next step is model fitting. Spot-On models the distribution of displacements for each subpopulation using Brownian motion under steady-state conditions without state transitions (full model description in section 1.3.4). Spot-On also accounts for localization errors (either user-defined or inferred from the SPT data). Crucially, Spot-On corrects for defocalization bias (Figure 1.18C) by explicitly calculating the probability that molecules move out-of-focus as a function of time and their diffusion constant. In fact, Spot-On uses the gradual loss of freely diffusing molecules over time as additional information to infer the diffusion constant and size of each subpopulation.

Spot-On considers either 2 or 3 subpopulations. For instance, TFs in nuclei can generally exist in both a chromatin-bound state characterized by slow diffusion and a freely diffusing state associated with rapid diffusion. In this case, a 2-state model is generally appropriate ('bound' vs. 'free'). Spot-On allows a user to choose their desired model and parameter ranges and then fits the model to the data. Using the previous example of TF dynamics, this allows the user to infer the bound fraction and the diffusion constants. Finally, once a user has finished fitting an appropriate model to their data, Spot-On allows easy download of publication-quality figures and relevant data (Figure 1.19; Full tutorial on the the Spot-On website).

1.3.4 Theory and implementation

1. **Spot-On model.**

   Spot-On implements and extends a kinetic modeling framework first described in(D. Mazza et al. 2012) and later extended in (Anders S. Hansen, Pustova, et al. 2017). Briefly, the model infers the diffusion constant and relative fractions of two or three subpopulations from the distribution of displacements (or histogram of displacements) computed at increasing lag time $(1\Delta\tau, 2\Delta\tau, \dots)$.

This is performed by fitting a semi-analytical model to the empirical histogram of displacements using non-linear least squares fitting. Defocalization is explicitly accounted for by modeling the fraction of particles that remain in focus over time as a function of their diffusion constant.

Mathematically, the evolution over time of a concentration of particles located at the origin as a Dirac delta function and which follows free diffusion in two dimensions with a diffusion constant D can be described by a propagator (also known as Green's function). Properly normalized, the probability of a particle starting at the origin ending up at a location $r = (x, y)$ after a time delay, $\Delta\tau$, is given by:

$$P(r, \Delta\tau) = N \frac{r}{2D\Delta\tau} e^{-\frac{r^2}{4D\Delta\tau}}$$

Here N is a normalization constant with units of length. Spot-On integrates this distribution over a small histogram bin window, Δr, to obtain a normalized distribution, the distribution of displacement lengths to compare to binned experimental data. For simplicity, we will therefore leave out N from subsequent expressions. Since experimental SPT data is subject to a significant mean localization error, σ, Spot-On also accounts for this (Matsuoka, Shibata, and Ueda 2009):

$$P(r, \Delta\tau) = \frac{r}{2\left(D\Delta\tau + \sigma^2\right)} e^{-\frac{r^2}{4\left(D\Delta\tau+\sigma^2\right)}}$$

Many proteins studied by SPT can generally exist in a quasi-immobile state (e.g. a chromatin-bound state in the case of transcription factors) and one or more mobile states. We will first consider the 2-state model. Under most conditions, state transitions can be ignored (Anders S. Hansen, Pustova, et al. 2017 and Figure 1.20). Thus, the steady-state 2-state model considered by Spot-On becomes:

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

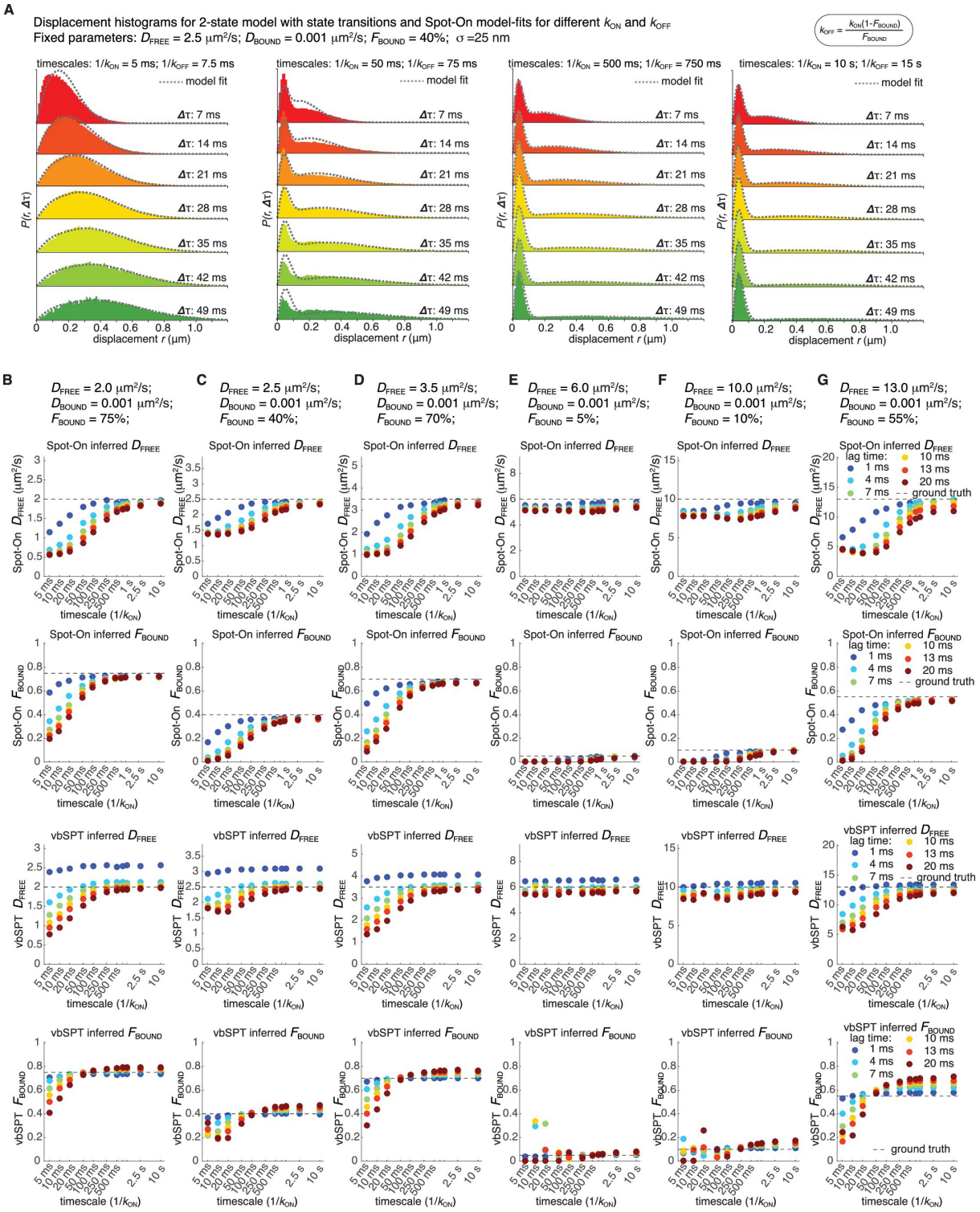


Figure 1.20: **Sensitivity of Spot-On to state changes and comparison with vbSPT.** For six different representative conditions (combinations of D_{FREE} and F_{BOUND}; D_{BOUND} = 0.001 μm²/s; σ = 25 nm), we simulated 100,000 trajectories using simSPT and included state transitions (e.g. transition from bound to free) considering six different lag time (1, 4, 7, 10, 13 and 20 ms) and k_{ON} values from 0.1 s^{-1} to 200 s^{-1} yielding a total of 396 simulations. The data were analyzed using Spot-On (all) as in Figure 1.23A-B. **(A)** For one example parameter set, (A) shows how the histogram of displacements and the goodness of the Spot-On model-fit changes as state transition go from more frequent than the frame rate (left) to very infrequent (right). **(B-G)**, *First row*: shows sensitivity of the Spot-On estimate of D_{FREE} to the timescale of state transitions. The values of D_{FREE} and F_{BOUND} are shown above the plot. **(B-G)**, *Second row*: shows sensitivity of the Spot-On estimate of F_{BOUND} to the timescale of state transitions. **(B-G)**, *Third row*: shows sensitivity of the vbSPT estimate of D_{FREE} to the timescale of state transitions. The values of D_{FREE} and F_{BOUND} are shown above the top plot. **(B-G)**, *Fourth row*: shows sensitivity of the vbSPT estimate of F_{BOUND} to the timescale of state transitions. As expected, since Spot-On ignores state transitions, the inference breaks down when the timescale of state transitions becomes comparable to the frame rate. Perhaps surprisingly, vbSPT also breaks down when state transitions are frequent despite explicitly modeling this in the Hidden Markov Model. Also as expected, a faster frame rate (e.g. 1 ms in dark blue) can support a faster state transition rate. Nevertheless, as long as the timescale of transitions is at least a few hundred milliseconds, Spot-On is not strongly affected. For comparison, the residence time of most mammalian transcription factors is tens of seconds. Source: `https://doi.org/10.7554/eLife.33125.016`

Here, the quasi-immobile subpopulation has diffusion constant, D_{BOUND}, and makes up a fraction, F_{BOUND}, whereas the freely diffusing subpopulation has diffusion constant, D_{FREE}, and makes up a fraction, $F_{FREE} = 1 - F_{BOUND}$. To account for defocalization bias (Figure 1.18C), Spot-On explicitly considers the probability of the freely diffusing subpopulation moving out of the axial detection range, Δz, during each time delay, $\Delta\tau$. This is important. For example, only $\sim 25\%$ of freely-diffusing molecules will remain in focus for at least five frames (assuming $\Delta\tau = 10$ ms; $\Delta z = 700$ nm; one gap allowed; $D = 5\mu m^2/s$), resulting in a 4-fold undercounting if uncorrected for. If we assume absorbing boundaries such that any molecule that contacts the edges of the axial detection range located at $z_{MAX} = \Delta z/2$ and $z_{MIN} = -\Delta z/2$ is permanently lost, the fraction of freely diffusing molecules with diffusion constant, D_{FREE}, that remain at time delay, $\Delta\tau$, is given by (Carslaw and Jaeger 1986; Kues and Kubitscheck 2002):

$$P_{remaining}(\Delta\tau) = \frac{1}{\Delta z}\int_{-\Delta z/2}^{\Delta z/2}\left\{1 - \sum_{n=0}^{\infty}(-1)^n\left[\mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z}{2} - z}{\sqrt{4D_{free}\Delta\tau}}\right) + \mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z}{2} + z}{\sqrt{4D_{free}\Delta\tau}}\right)\right]\right\}dz$$

However, this analytical expression overestimates the fraction lost since there is a significant probability that a molecule that briefly contacted or exceeded the boundary re-enters the axial detection range. The re-entry probability depends on the number of gaps (g) allowed in the tracking, $\Delta\tau$, and Δz and can be approximately accounted for by considering a corrected axial detection range, δz_{corr}, larger than Δz: $\Delta z_{corr} > \Delta z$:

$$\Delta z_{corr}(\Delta z, \Delta\tau, D) = \Delta z + a(\Delta z, \Delta\tau)\sqrt{D} + b(\Delta z, \Delta\tau)$$

Although Δz_{corr} depend on the number of gaps (g) allowed in the tracking, we will leave it out for simplicity in the following. We determined the coefficients a and b from Monte Carlo simulations. For a given diffusion constant, D, 50,000 molecules were randomly placed one-dimensionally along the z-axis drawn from a uniform distribution from $z_{MIN} = -\Delta z/2$ to $z_{MAX} = \Delta z/2$. Next, using a time-step $\Delta\tau$, one-dimensional Brownian diffusion was simulated along the z-axis using the Euler-Maruyama scheme. For time delays from $1\Delta\tau$ to $15\Delta\tau$, the fraction of molecules that were lost was calculated in the range of $D = [1; 12]\mu m^2/s$. $a(\Delta z, \Delta\tau, g)$ and $b(\Delta z, \Delta\tau, g)$ were then estimated through least-squares fitting of $P_{remaining}(\Delta\tau, \Delta z_{corr}, D)$ to the simulated fraction remaining. The process was repeated over a grid of plausible values of $(\Delta z, \Delta\tau, g)$ to derive a grid of 134,865 (a, b) parameter pairs. This pre-calculated library of (a, b) parameters enables Spot-On to perform model fitting on nearly any SPT dataset with minimal overhead.

Thus, the 2-state model Spot-On uses for kinetic modeling of SPT data is given by:

$$P(r, \Delta\tau) = F_{bound}\frac{r}{2\left(D_{bound}\Delta\tau + \sigma^2\right)}e^{-\frac{r^2}{4(D_{bound}\Delta\tau + \sigma^2)}} \tag{1.1}$$

$$+Z_{corr}(\Delta\tau)\left(1 - F_{bound}\right)\frac{r}{2\left(D_{free}\Delta\tau + \sigma^2\right)}e^{-\frac{r^2}{4\left(D_{free}\Delta\tau + \sigma^2\right)}} \tag{1.2}$$

where:

$$Z_{corr}(\Delta\tau) = \frac{1}{\Delta z}\int_{-\Delta z/2}^{\Delta z/2}\left\{1 - \sum_{n=0}^{\infty}(-1)^n\left[\mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z_{corr}}{2} - z}{\sqrt{4D_{free}\Delta\tau}}\right) + \mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z_{corr}}{2} + z}{\sqrt{4D_{free}\Delta\tau}}\right)\right]\right\}dz$$

Having derived the 2-state model, generalization to a 3-state model with 1 bound and 2 diffusive states is straightforward. If the three subpopulations have diffusion constants D_{BOUND}, D_{SLOW},

D_{FAST}, and fractions F_{BOUND}, F_{SLOW}, F_{FAST}, such that $F_{BOUND} + F_{SLOW} + F_{FAST} = 1$, then the 3-state model considered by Spot-On becomes:

$$P_3(r,\Delta\tau) = F_{bound}\frac{r}{2\left(D_{bound}\Delta\tau\sigma^2\right)}e^{-\frac{r^2}{4\left(D_{bound}\Delta\tau+\sigma^2\right)}} \tag{1.3}$$

$$+F_{slow}\frac{r}{2\left(D_{slow}\Delta\tau+\sigma^2\right)}e^{-\frac{r^2}{4\left(D_{slow}\Delta\tau+\sigma^2\right)}} \tag{1.4}$$

$$+\left(1-F_{bound}-F_{slow}\right)\frac{r}{2\left(D_{fast}\Delta\tau+\sigma^2\right)}e^{-\frac{r^2}{4\left(D_{fast}\Delta\tau+\sigma^2\right)}} \tag{1.5}$$

Where $Z_{CORR}(\Delta\tau, \Delta z_{corr}, D)$ is as described above.

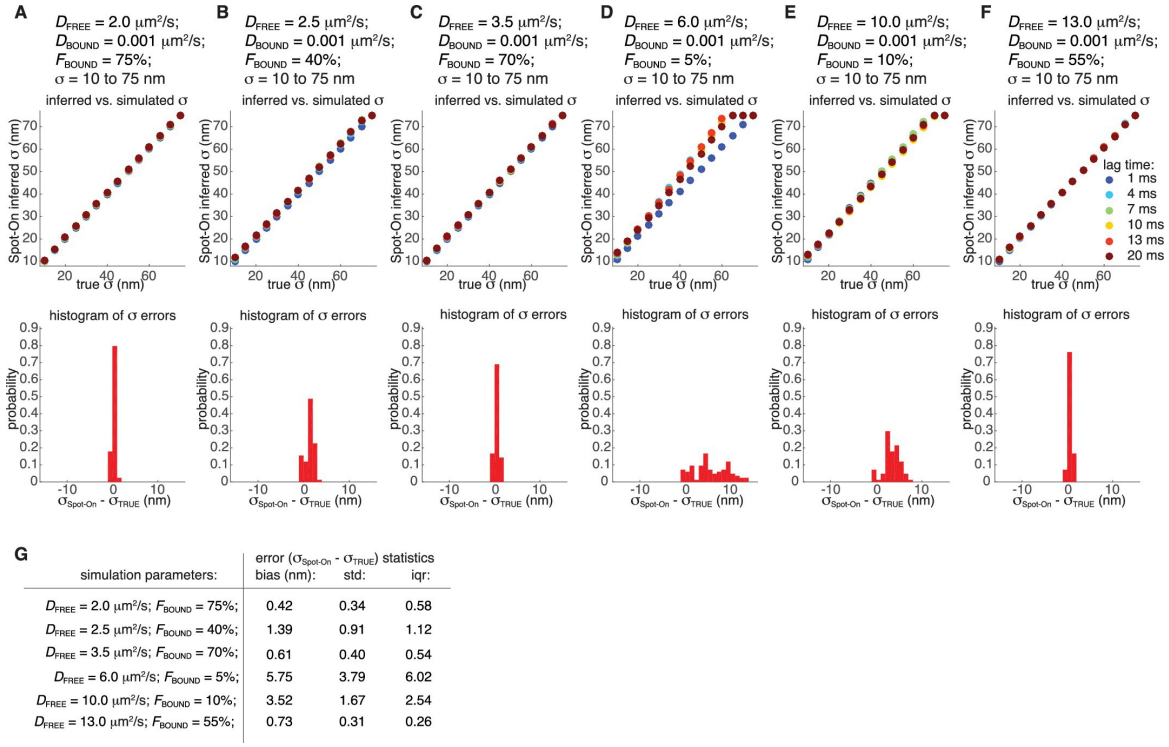
2. **Numerical implementation of models in Spot-On.**

   Spot-On calculates the empirical histogram of displacements based on a user-defined bin width. Spot-On allows the user to choose between PDF- and CDF-fitting of the kinetic model to the empirical displacement distributions; CDF-fitting is generally most accurate for smaller datasets and the two are similar for large datasets (Anders S. Hansen, Woringer, et al. 2018; Figure 3—figure supplement 9). The integral in $\mathrm{Z_{CORR}}(\Delta\tau,\Delta\ \mathrm{z_{corr}})$ was numerically evaluated using the midpoint method over 200 points and the terms of the series computed until the term falls below a threshold of 10^{-10}. Model fitting and parameter optimization was performed using a non-linear least squares algorithm (Levenberg-Marquardt). Random initial parameter guesses are drawn uniformly from the user-specified parameter range. The optimization is then repeated several times with different initialization parameters to avoid local minima. Spot-On constrains each fraction to be between 0 and 1 and for the sum of the fractions to equal 1.

3. **Theoretical characteristics and limitations of the model.**

   Although Spot-On performs well on both experimental and simulated SPT data, the model implemented by Spot-On has several limitations. First, the kinetic model assumes diffusion to be ideal Brownian motion, even though it is widely acknowledged that the motion of most proteins inside a cell shows some degree of anomalous diffusion. We propose an extension of Spot-On to take into account anomalous diffusion in section IV.2. Nevertheless, 1.26G–H and (Anders S. Hansen, Woringer, et al. 2018; Figure 4—supplement 2) show that the parameter inference for experimental data of proteins presenting various degrees of anomalous diffusion is quite robust.

   Second, Spot-On models the localization error as the static mean localization error and this feature can be used to infer the actual localization error from the data. However, the localization error is affected both by the position of the particle with respect to the focal plane (Lindén, Ćurić, Amselem, et al. 2017) and by motion blur (Deschout, Neyts, and Braeckmans 2012). Even though a high signal-to-background ratio and fast framerate/stroboscopic illumination help to mitigate these disparities, it is likely that the localization error of fast moving particles will be higher than the bound/slow-moving particles. In that case, one would expect Spot-On to infer a localization error that is the weighted mean of the 'bound/static' localization error and the 'free' localization error. However, in many situations $D_{free}\Delta\tau >> \sigma^2$ (even assuming a $2\mu m^2/s$ particle imaged at a 5 ms framerate with a $\sim$ 30 nm localization error, there is still an order of magnitude difference between the two terms). As a consequence, the estimate of σ reflects the static localization error (that is, the localization error of the bound fraction), and the localization error estimate becomes less reliable if the bound fraction is very small (Figure 1.21).



G

simulation parameters:	error ($σ_{Spot-On}$ - $σ_{TRUE}$) statistics bias (nm):	std:	iqr:
D_{FREE} = 2.0 μm²/s; F_{BOUND} = 75%;	0.42	0.34	0.58
D_{FREE} = 2.5 μm²/s; F_{BOUND} = 40%;	1.39	0.91	1.12
D_{FREE} = 3.5 μm²/s; F_{BOUND} = 70%;	0.61	0.40	0.54
D_{FREE} = 6.0 μm²/s; F_{BOUND} = 5%;	5.75	3.79	6.02
D_{FREE} = 10.0 μm²/s; F_{BOUND} = 10%;	3.52	1.67	2.54
D_{FREE} = 13.0 μm²/s; F_{BOUND} = 55%;	0.73	0.31	0.26

Figure 1.21: **Robustness of localization error estimates from Spot-On.** For six different representative conditions (combinations of D_{FREE} and F_{BOUND}; D_{BOUND} = 0.001 μm²/s), we simulated 100,000 trajectories using simSPT keeping everything as in Figure 1.23A–B except varying the localization error (σ) from 10 nm to 75 nm in 5 nm steps and considering six different lag time (1, 4, 7, 10, 13 and 20 ms) yielding a total of 504 simulations. The data were analyzed using Spot-On (all) as in Figure 1.23A–B except here the localization error was inferred from the fitting. **(A-F), top row**: show how well the Spot-On inferred the localization error vs. the simulated localization error and the lag times are color coded. The values of D_{FREE} and F_{BOUND} are shown above the plot. **(A-F), bottom row**: histograms showing the distribution of errors in the localization error estimate across all lag time and σ-values for a given combinations of D_{FREE} and F_{BOUND}. **(G)** Table showing summary statistics from the fitting in (A-F). We note that in all cases where the bound fraction is significant (>10%), Spot-On robustly infers the localization error (mean error below 1.5 nm), whereas in cases where the bound fraction is small (10% or below), the localization error estimate becomes less robust (mean error $\sim$ 3–6 nm). This is because Spot-On can most reliably use how the displacement distribution of the bound fraction changes over time to infer the localization error. Source: `https://doi.org/10.7554/eLife.33125.017`

Third, following (Kues and Kubitscheck 2002) the axial detection profile is assumed to be a step function, which is an approximation. However, all simulations here were performed using a detection profile with Gaussian edges (Anders S. Hansen, Woringer, et al. 2018 Figure 3—supplement 1) and as shown in Figure 1.23A–B Spot-On still works quite well and moreover is relatively robust to slight mismatches in the axial detection range (Anders S. Hansen, Woringer, et al. 2018 Figure 3—figure supplement 7).

Fourth, unlike the original implementation by (D. Mazza et al. 2012), Spot-On ignores state transitions. This reduces the number of fitted parameters and simplifies the generalization to more than two states, but as shown in Figure 1.20 it also causes the parameter inference to fail unless the timescale of state changes is at least 10–50 times longer than the frame rate. Thus, in cases where a molecule is known to exhibit state changes on a time-scale of tens to a few hundreds of milliseconds, Spot-On may not be appropriate.

Fifth and finally, Spot-On ignores correlations between adjacent displacements, although taking such information into account can potentially improve the parameter inference (Vestergaard, Blainey, and Flyvbjerg 2014).

1.3.5 Validation of Spot-On using simulated SPT data and comparison to other methods

We first evaluated whether Spot-On could accurately infer subpopulations (Figure 1.18D) and successfully account for known biases (Figure 1.18C) using simulated data. We compared Spot-On to a popular alternative approach of first fitting the mean square displacement (MSD) of individual trajectories of a minimum length and then fitting the distribution of estimated diffusion constants (we refer to this as 'MSD_i') as well as a sophisticated Hidden-Markov Model-based Bayesian inference method (vbSPT) (Persson et al. 2013). Since most SPT data is collected using highly inclined illumination (Tokunaga, Imamoto, and Sakata-Sogawa 2008) (HiLo), we simulated TF binding and diffusion dynamics (2-state model: 'bound vs. free') confined inside a $4\mu m$ radius mammalian nucleus under realistic HiLo SPT experimental settings subject to a 25 nm localization error (Figure 1.22). We considered the effect of the exposure time (1 ms, 4 ms, 7 ms, 13 ms, 20 ms), the free diffusion constant (from $0.5\mu m^2/s$ to $14.5\mu m^2/s$ in $0.5\mu m^2/s$ increments) and the bound fraction (from 0% to 95% in 5% increments) yielding a total of 3480 different conditions that span the full range of biologically plausible dynamics (Figure **??** and Anders S. Hansen, Woringer, et al. 2018 Figure 3—figure 3; Appendix 1).

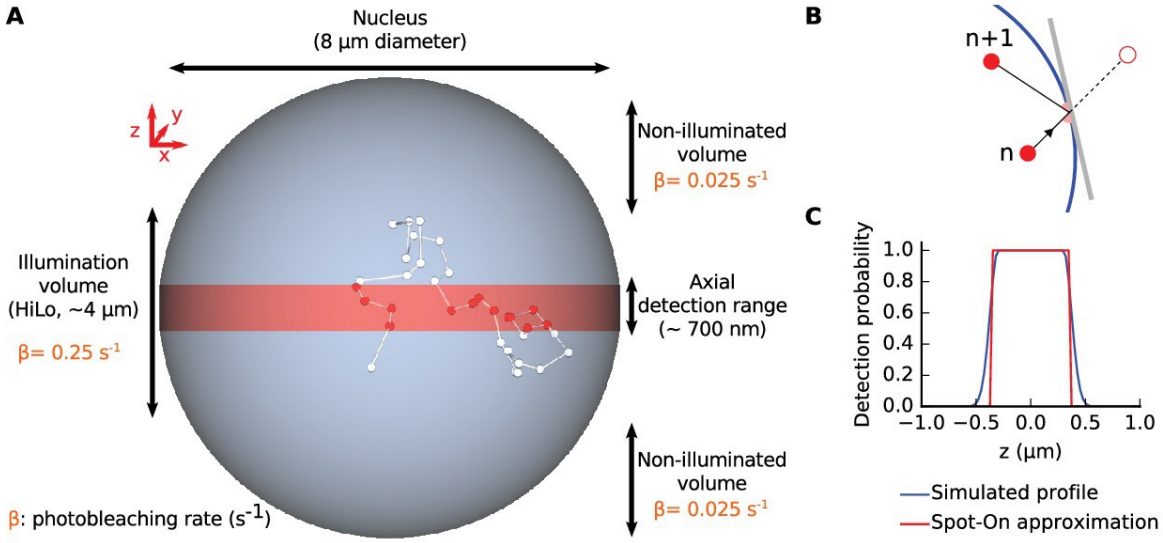


Figure 1.22: **Overview of SPT simulations. (A)** Trajectories were simulated in a confined volume: a 'nucleus' of 8 µm diameter, in which molecules are photoactivated at random and photobleach when located within the HiLo volume (a ∼ 4 µm thick slice). Molecules are detected when they are within the axial detection range of the objective (∼ 700 nm). **(B)** confinement within the nucleus was achieved by specular reflections against the nuclear envelope: a particle bumping on the nuclear envelope is ballistically reflected inside. **(C)** axial detection profile used for the simulation (*blue*): flat-top Gaussian with 600 nm plateau and 100 nm FWHM for the Gaussian edges. (*red*): approximated axial detection profile assumed by Spot-On (step function with 700 nm width). Source: `https://doi.org/10.7554/eLife.33125.007`

Spot-On accurately inferred subpopulation sizes with minimal error (Figure 1.23A–B, Table 1.2), but slightly underestimated the diffusion constant (-4.8%; Figure 1.23B; Table 1.2). However, this underestimate was due to particle confinement inside the nucleus: Spot-On correctly inferred the diffusion constant when the confinement was relaxed (Anders S. Hansen, Woringer, et al. 2018 Figure 3—supplement 4; $20\mu m$ nuclear radius instead of $4\mu m$). This emphasizes that diffusion constants

measured by SPT inside cells should be viewed as apparent diffusion constants. In contrast, the MSD_i method failed under most conditions regardless of whether all trajectories were used (MSD_i (all)) or a fitting filter applied (MSD_i ($R^2 > 0.8$); Figure 3A–B; Table 1). vbSPT performed almost as well as Spot-On for slow-diffusing proteins, but showed larger deviations for fast-diffusing proteins (Figure **??** and Anders S. Hansen, Woringer, et al. 2018 Figure 3—figure 3).

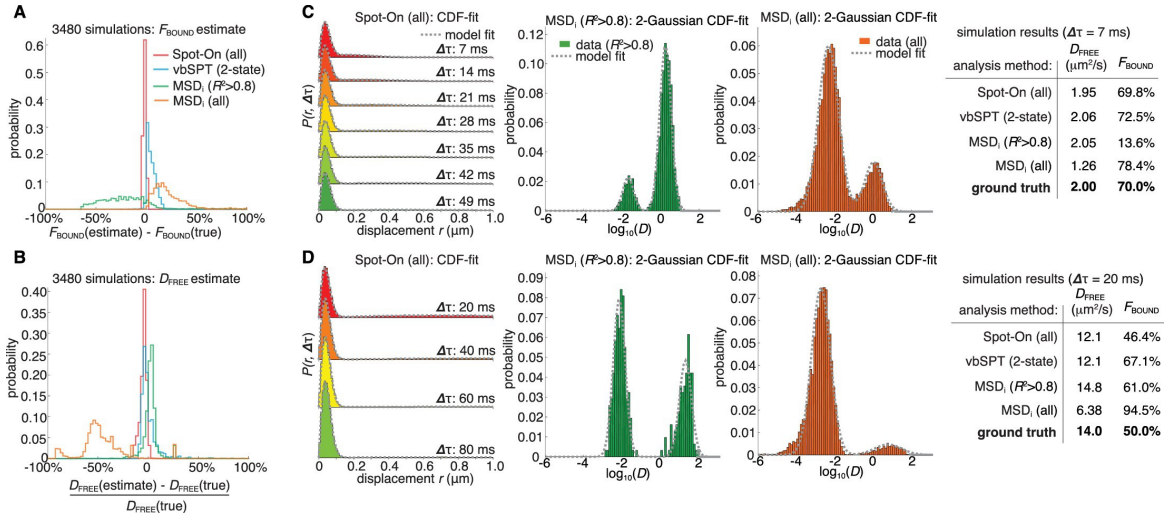


Figure 1.23: **Validation of Spot-On using simulations and comparisons to other methods. (A–B)** Simulation results. Experimentally realistic SPT data was simulated inside a spherical mammalian nucleus with a radius of 4 µm subject to highly-inclined and laminated optical sheet illumination (Tokunaga, Imamoto, and Sakata-Sogawa 2008) (HiLo) of thickness 4 µm illuminating the center of the nucleus. The axial detection window was 700 nm with Gaussian edges and particles were subject to a 25 nm localization error in all three dimensions. Photobleaching corresponded to a mean trajectory length of 4 frames inside the HiLo sheet and 40 outside. 3480 experiments were simulated with parameters of $D_{FREE} = [0.5; 14.5]$ in steps of $0.5\mu m^2/s$ and $F_{BOUND} = [0; 95\%]$ in steps of 5% and the frame rate correspond to $\Delta\tau = [1, 4, 7, 10, 13, 20]$ ms. Each experiment was then fitted using Spot-On, using vbSPT (maximum of 2 states allowed) (persson$_{\text{extracting2013}}$ et al., 2013), MSD_i using all trajectories of at least five frames (MSD_i (all)) or MSD_i using all trajectories of at least five frames where the MSD-curvefit showed at least $R^2 > 0.8$ (MSD_i ($R^2 > 0.8$)). **(A)** shows the distribution of absolute errors in the F_{BOUND}–estimate and **(B)** shows the distribution of relative errors in the D_{FREE}–estimate. **(C)** Single simulation example with $D_{FREE} = 2.0\mu m^2/s$; $F_{BOUND} = 70\%$; 7 ms per frame. The table on the right uses numbers from CDF-fitting, but for simplicity the fits to the histograms (PDF) are shown in the three plots. **(D)** Single simulation example with $D_{FREE} = 14.0\mu m^2/s$; $F_{BOUND} = 50\%$; 20 ms per frame. Full details on how SPT data was simulated and analyzed with the different methods is given in Anders S. Hansen, Woringer, et al. 2018; Appendix 1. Source: `https://doi.org/10.7554/eLife.33125.006`

Table 1.2: **Summary of simulation results and comparison of methods**. The table shows the bias (mean error), 'std' (standard deviation) and 'iqr' (inter-quartile range: difference between the 75th and 25th percentile) for each method for all 3480 simulations. The left column shows the relative bias/std/iqr for the D_{FREE}-estimate and the right column shows the absolute bias/std/iqr for the F_{BOUND}-estimate. Source: `https://doi.org/10.7554/eLife.33125.019`

Analysis method	D_{FREE}			F_{BOUND}		
	bias	std	iqr	bias	std	iqr
Spot-On (all)	-4.8%	3.3%	3.5%	-1.7%	1.2%	1.8%
vbSPT (2-state)	0.8%	12.5%	6.8%	5.0%	4.6%	6.1%
MSD_i ($R^2 > 0.8$)	8.0%	28.5%	4.9%	-20.6%	26.4%	32.1%
MSD_i (all)	-39.6%	41.8%	19.0%	22.0%	15.8%	17.8%

To illustrate how the methods could give such divergent results when run on the same SPT data, we considered two example simulations (Figure 1.23C–D; more examples in Anders S. Hansen, Woringer, et al. 2018; Figure 3—figure supplement 3). First, we considered a mostly bound and relatively slow diffusion case (D_{FREE} : $2.0\mu m^2/s$; F_{BOUND}: 70%; $\Delta\tau$: 7 ms; Figure 3C). Spot-On and vbSPT accurately inferred both D_{FREE} and F_{BOUND}. In contrast, MSD_i ($R^2 > 0.8$) greatly underestimated F_{BOUND} (13.6% vs. 70%), whereas MSD_i (all) slightly overestimated F_{BOUND}. Since MSD_i-based methods apply two thresholds (first, minimum trajectory length: here five frames; second, filtering based on R^2) in many cases less than 5% of all trajectories passed these thresholds and this example illustrate how sensitive MSD_i-based methods are to these thresholds. Note that although we show the fits to the probability density function since this is more intuitive (PDF; histogram), we performed the fitting to the cumulative distribution function (CDF). Second, we considered an example with a slow frame rate and fast diffusion, such that the free population rapidly moves out-of-focus (D_{FREE} : $14.0\mu m^2/s$; F_{BOUND}: 50%; $\Delta\tau$: 20 ms; Figure 3D). Spot-On again accurately inferred F_{BOUND}, and slightly underestimated D_{FREE} due to high nuclear confinement (Anders S. Hansen, Woringer, et al. 2018; Figure 3—supplement 4). Although vbSPT generally performed well, because it does not correct for defocalization bias (vbSPT was developed for bacteria, where defocalization bias is minimal), vbSPT strongly overestimated F_{BOUND} in this case (Figure 1.23D). Consistent with this, Spot-On without defocalization-bias correction also strongly overestimates the bound fraction (Anders S. Hansen, Woringer, et al. 2018; Figure 3—supplement 5). We conclude that correcting for defocalization bias is critical. The MSD_i-based methods again gave divergent results despite seemingly fitting the data well. Thus, a good fit to a histogram of log(D) does not necessarily imply that the inferred D_{FREE} and F_{BOUND} are accurate. A full discussion and comparison of the methods is given in (Anders S. Hansen, Woringer, et al. 2018; Appendix 1). Finally, we extended this analysis of simulated SPT data to three states (one 'bound', two 'free' states) and compared Spot-On and vbSPT. Spot-On again accurately inferred both the diffusion constants and subpopulation fractions of each population and slightly outperformed vbSPT (Figure 1.24).

Having established that Spot-On is accurate, we next tested whether it was also robust. Spot-On's ability to infer D_{FREE} and F_{BOUND} was robust to misestimates of the axial detection range of $\sim$ 100–200 nm (Anders S. Hansen, Woringer, et al. 2018; Figure 3—supplement 7), was minimally affected by the number of timepoints considered and fitting parameters (Anders S. Hansen, Woringer, et al. 2018; Figure 3—supplements 8–9; see also Appendix 2 for parameter considerations) and was not strongly affected by state changes (e.g. binding or unbinding) provided the time-scale of state changes is significantly longer than the frame rate (Figure 1.20). Moreover, Spot-On inferred the localization error with nanometer precision provided that a significant bound fraction is present (Figure 1.21). Finally, we sub-sampled the data sets and found that just $\sim$ 3000 short trajectories (mean length $\sim$ 3–4 frames) were sufficient for Spot-On to reliably infer the underlying dynamics (Figure 1.25). We conclude that Spot-On is robust.

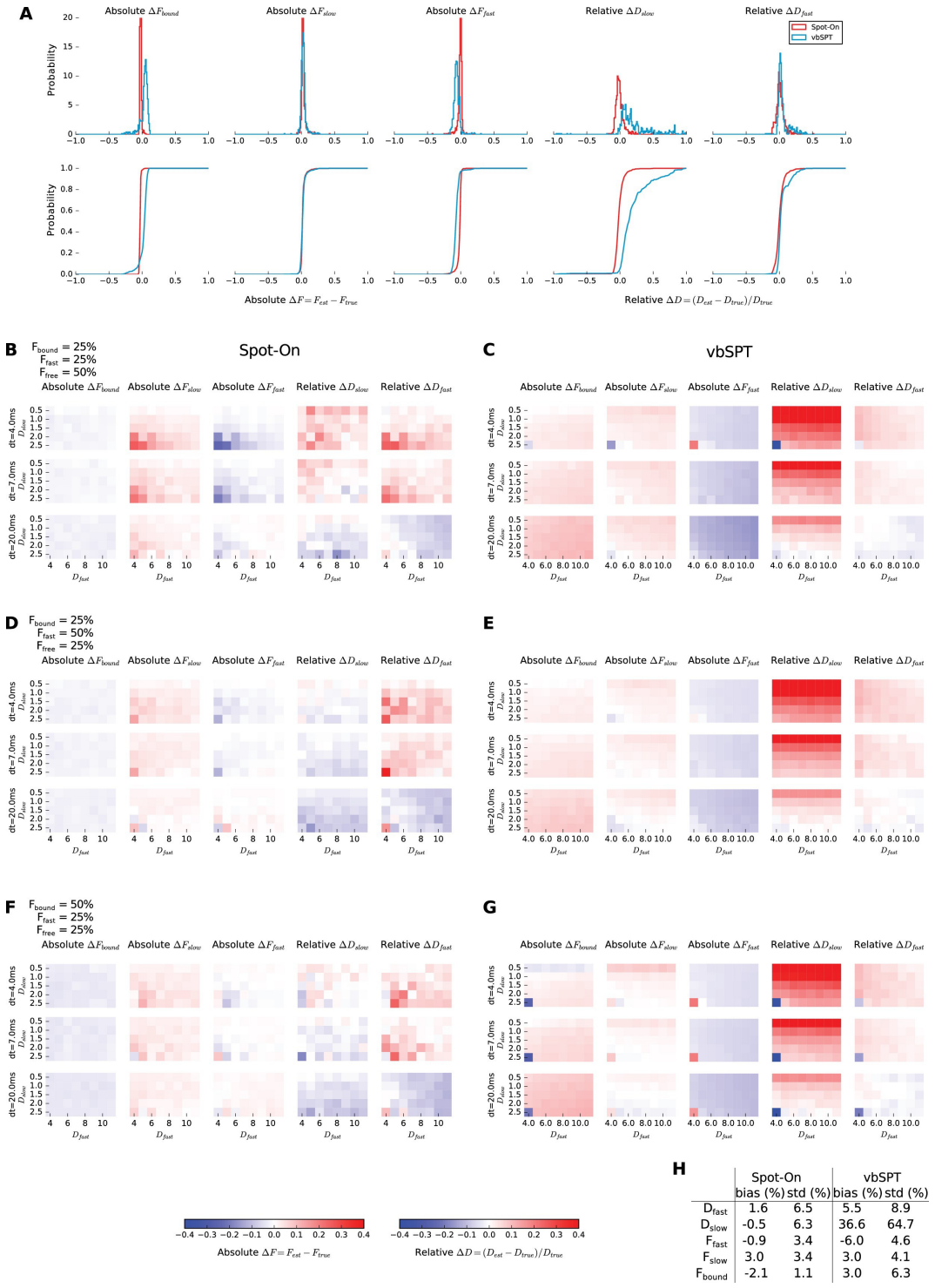


Figure 1.24: **Evaluation of the 3-states model.** Trajectories were simulated using simSPT for a 3-state model. Three representative fractions were picked and for each of them, one state was always bound (D_{BOUND} = 0.001 µm²/s) and the two other states were varied (0.5–11 µm²/s), together with the framerate (1–20 ms), yielding 720 conditions. The simulations were then either fitted with Spot-On or vbSPT constrained to infer up to three states. **(A)** Distribution of the error of five of the inferred parameters (D_{SLOW}, D_{FAST}, F_{BOUND}, F_{SLOW}, F_{FAST}) with respect to ground truth for Spot-On (*red*) and vbSPT (*blue*). The top row shows the distribution and the bottom row the cumulative distribution. **(B-G)** For each of the three fractions configurations (25/25/50, 25/50/25, 50/25/25%, for B-C, D-E, F-G, respectively), detailed error on five inferred parameters (columns) for different frame rates (rows) and various D_{SLOW} and D_{FAST} (rows and columns of the matrix, respectively). **(H)** summary table showing the mean error (bias) and standard deviation over all the simulations. Source: `https://doi.org/10.7554/eLife.33125.012`

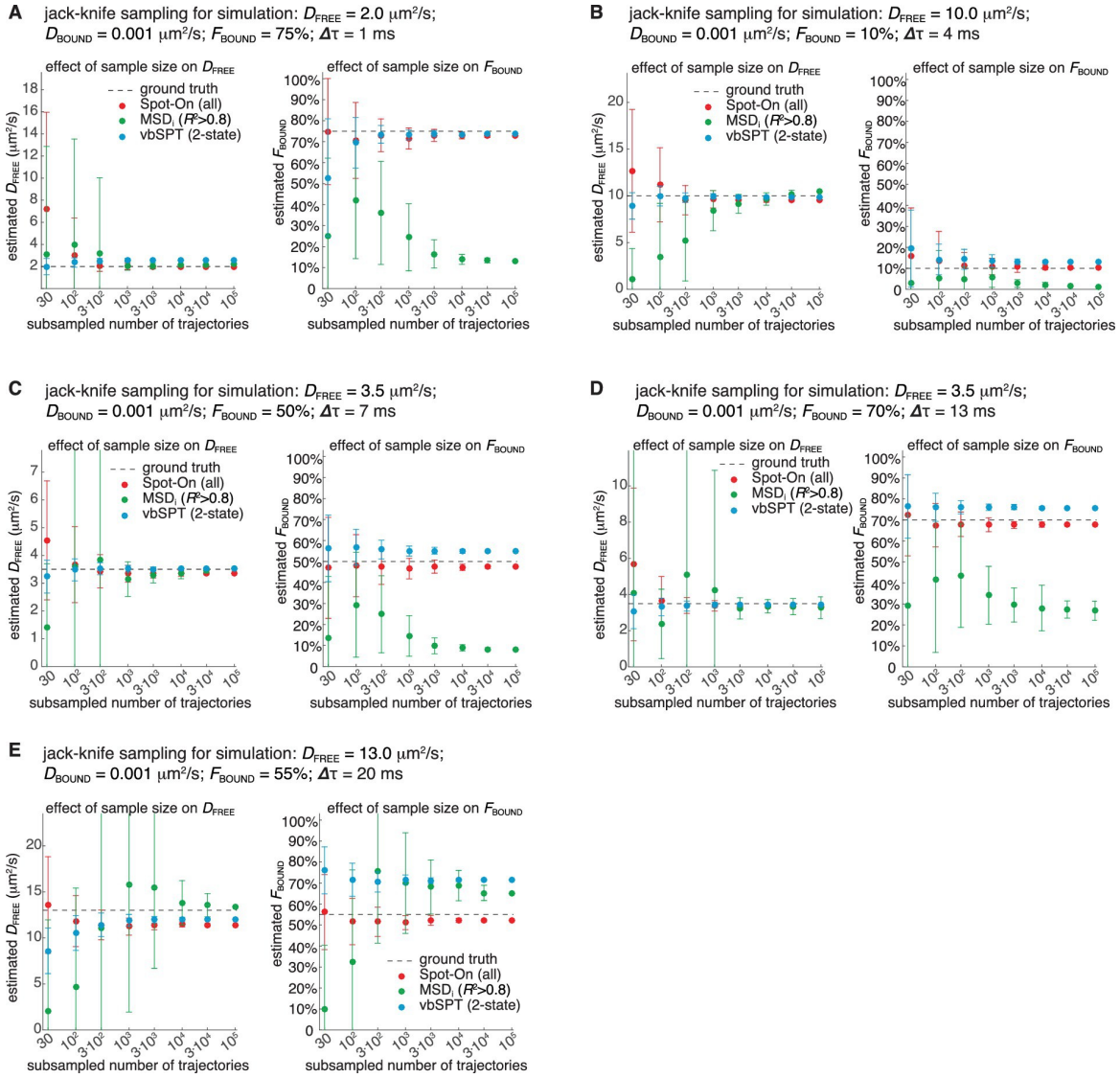


Figure 1.25: **Sensitivity of Spot-On, vbSPT and MSD_i (R2 >0.8) to sample size. (A)** Jack-knife data sampling for simulation with D_{FREE} = 2.0 μm^2/s; F_{BOUND} = 75%; 1 ms per frame. Simulated data (inside a 4 μm radius nucleus) was used. 100,000 trajectories with a mean photo-bleaching life-time of 4 frames were simulated and then subsampled 50 times without replacement. Sample sizes of either 30, 100, 300, 1,000, 3,000, 10,000, 30,000 or 100,000 trajectories were then fit using Spot-On (all), vbSPT (2-state model) or MSD_i (R^2 >0.8) as described in the analysis of simulations section. Error bars show standard deviation among the 50 sub-samplings. We note that occasionally, no more than ∼ 5% of the time in the case of 30 trajectories, not a single trajectory of sufficient length for Spot-On or MSD_i (R^2 >0.8) was found. In these cases, we re-sampled to obtain at least one trajectory of sufficient length. *Left plot* shows effect of sample size on the D_{FREE}–estimate. *Right plot* shows effect of sample size on the F_{BOUND}–estimate. The dashed line shows the ground truth used to simulate the SPT data. **(B)** Jack-knife data sampling for simulation with D_{FREE} = 10.0 μm^2/s; F_{BOUND} = 10%; 4 ms per frame. Everything else is as described in (A). **(C)** Jack-knife data sampling for simulation with D_{FREE} = 3.5 μm^2/s; F_{BOUND} = 50%; 7 ms per frame. Everything else is as described in (A). **(D)** Jack-knife data sampling for simulation with D_{FREE} = 3.5 μm^2/s; F_{BOUND} = 70%; 13 ms per frame. Everything else is as described in (A). **(E)** Jack-knife data sampling for simulation with D_{FREE} = 13.0 μm^2/s; F_{BOUND} = 55%; 20 ms per frame. Everything else is as described in (A). `https://doi.org/10.7554/eLife.33125.018`

Taken together, this analysis of simulated SPT data suggests that Spot-On successfully overcomes

defocalization and analysis method biases (Figure 1.18C–D), accurately and robustly estimates sub-populations and diffusion constants across a wide range of dynamics and, finally, outperforms other methods.

1.3.6 Validations on experimental data

Having validated Spot-On on simulated data, which is not subject to experimental biases (Figure 1.18A–B), we next sought to evaluate Spot-On on experimental data. To generate SPT data with minimal acquisition bias we performed stroboscopic photo-activation SPT (spaSPT; Figure 1.26A), which integrates previously and separately published ideas to minimize experimental biases. First, spaSPT minimizes motion-blurring, which is caused by particle movement during the camera exposure time (Figure 1.18A), by using stroboscopic excitation (J. Elf, G.-W. Li, and X. S. Xie 2007; Frost, H. E. Lu, and Blanpied 2012). We found that the bright and photo-stable dyes PA-JF_{549} and PA-JF_{646} (Jonathan B Grimm, Brian P English, J. Chen, et al. 2015) in combination with the HaloTag ('Halo') labeling strategy made it possible to achieve a signal-to-background ratio greater than 5 with just 1 ms excitation pulses, thus providing a good compromise between minimal motion-blurring and high signal (Figure 1.26B). Second, spaSPT minimizes tracking errors (Figure 1.18B) by using photo-activation (Figure 1.26A) (Jonathan B Grimm, Brian P English, J. Chen, et al. 2015; Manley et al. 2008). Tracking errors are generally caused by high particles densities. Photo-activation allows tracking at extremely low densities ($\leq$ 1 molecule per nucleus per frame) and thereby minimizes tracking errors (Izeddin, Récamier, et al. 2014), whilst at the same time generating thousands of trajectories. To consider the full spectrum of nuclear protein dynamics, we studied histone H2B-Halo (overwhelmingly bound; fast diffusion; Figure 1.26C), Halo-CTCF (Anders S. Hansen, Pustova, et al. 2017) (largely bound; slow diffusion; Figure 1.26D) and Halo-NLS (overwhelmingly free; very fast diffusion; Figure 1.26F) in human U2OS cells and Halo-Sox2 (Teves, An, Anders S Hansen, et al. 2016) (largely free; intermediate diffusion; Figure 1.26E) in mouse embryonic stem cells (mESCs). We labeled Halo-tagged proteins in live cells with the HaloTag ligands PA-JF_{549} or PA-JF_{646} (Jonathan B Grimm, Brian P English, J. Chen, et al. 2015) and performed spaSPT using HiLo illumination (Anders S. Hansen, Woringer, et al. 2018; Video 2). To generate a large dataset to comprehensively test Spot-On, we performed 1064 spaSPT experiments across 60 different conditions.

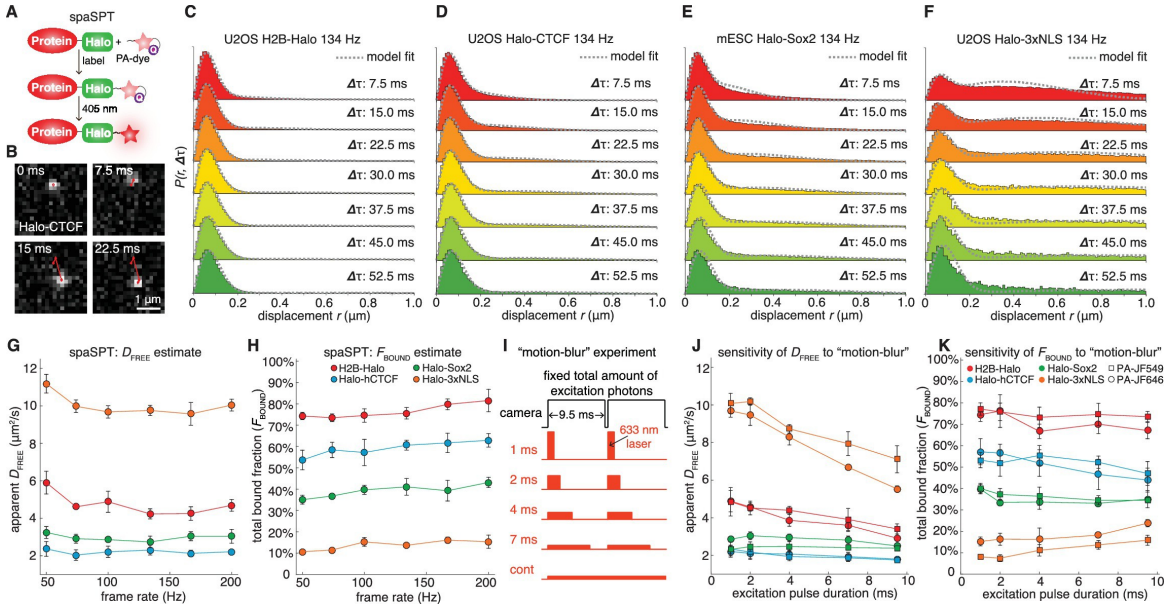


Figure 1.26: **Overview of spaSPT and experimental results. (A)** spaSPT. HaloTag-labeling with UV (405 nm) photo-activatable dyes enable spaSPT. spaSPT minimizes tracking errors through photo-activation which maintains low densities. **(B)** Example data. Raw spaSPT images for Halo-CTCF tracked in human U2OS cells at 134 Hz (1 ms stroboscopic 633 nm excitation of JF_{646}). **(C–F)** Histograms of displacements for multiple $\Delta\tau$ of histone H2B-Halo in U2OS cells **(C)**, Halo-CTCF in U2OS cells **(D)**, Halo-Sox2 in mES cells **(E)** and Halo-3xNLS in U2OS cells **(F)**. **(G–H)** Effect of frame-rate on D_{FREE} and F_{BOUND}. spaSPT was performed at 200 Hz, 167 Hz, 134 Hz, 100 Hz, 74 Hz and 50 Hz using the 4 cell lines and the data fit using Spot-On and a 2-state model. Each experiment on each cell line was performed in four replicates on different days and $\sim$ 5 cells imaged each day. **(I)** Motion-blur experiment. To investigate the effect of 'motion-blurring', the total number of excitation photons was kept constant, but delivered during pulses of duration 1, 2, 4, 7 ms or continuous (cont) illumination. **(J–K)** Effect of motion-blurring on D_{FREE} and F_{BOUND}. spaSPT data was recorded at 100 Hz and 2-state model-fitting performed with Spot-On. The inferred D_{FREE} **(J)** and F_{BOUND} **(K)** were plotted as a function of excitation pulse duration. Each experiment on each cell line was performed in four replicates on different days and $\sim$ 5 cells imaged each day. Error bars show standard deviation between replicates. Source: `https://doi.org/10.7554/eLife.33125.020`

1. *Validation of Spot-On using spaSPT data at different frame rates*

   First, we studied whether Spot-On could consistently infer subpopulations over a wide range of frame rates. We experimentally determined the axial detection range to be $\sim$ 700 nm (Anders S. Hansen, Woringer, et al. 2018; Figure 4—figure supplement 1) and performed spaSPT at 200 Hz, 167 Hz, 134 Hz, 100 Hz, 74 Hz and 50 Hz using the four cell lines. Spot-On consistently inferred the diffusion constant (Figure 1.26G) and total bound fraction across the wide range of frame rates (Figure 1.26H). This is notable since all four proteins exhibit apparent anomalous diffusion (Anders S. Hansen, Woringer, et al. 2018; Figure 4—supplement 2) and this demonstrates that Spot-On is also robust to anomalous diffusion despite modeling Brownian motion. While the ground-truth is unknown when considering experiments, Spot-On gave biologically reasonable results: histone H2B was overwhelmingly bound and free Halo-3xNLS was overwhelmingly unbound (comparison with vbSPT: Anders S. Hansen, Woringer, et al. 2018; Figure 4—figure supplement 3). These results provide additional validation for the bias corrections implemented in Spot-On. We also note that although Spot-On was validated on spaSPT data, SPT data with non-photoactivatable dyes is also suitable for Spot-On analysis provided that the density is sufficiently low to minimize tracking errors (see also Anders S. Hansen, Woringer, et al. 2018; Appendix 3: "Which datasets are appropriate for Spot-On?"). Finally, we demonstrated above

that just $\sim$ 3000 short trajectories (mean length $\sim$ 3–4 frames) were sufficient for Spot-On to accurately infer D_{FREE} and F_{BOUND} (Figure 1.25). Here we obtain well above 3000 trajectories per cell even at $\sim$ 1 localization/frame. More generally, with spaSPT this should be generally achievable for all but the most lowly expressed nuclear proteins. Thus, this now makes it possible to study biological cell-to-cell variability in TF dynamics.

2. *Effect of motion-blur bias on parameter estimates*

   Having validated Spot-On on experimental SPT data, we next applied Spot-On to estimate the effect of motion-blurring on the estimation of subpopulations. As mentioned, since most localization algorithms (Chenouard et al. 2014; Sergé et al. 2008) achieve super-resolution through PSF-fitting, this may cause motion-blurred molecules to be undersampled, resulting in a bias towards slow-moving molecules (Figure 1.18A). We estimated the extent of the bias by imaging the four cell lines at 100 Hz and keeping the total number of excitation photons constant, but varying the excitation pulse duration (1 ms, 2 ms, 4 ms, 7 ms, constant; Figure 1.26I). For generality, we performed these experiments using both PA-JF_{549} and PA-JF_{646} dyes (Jonathan B Grimm, Brian P English, J. Chen, et al. 2015). We used Spot-On to fit the data and plotted the apparent free diffusion constant (Figure 1.26J) and apparent total bound fraction (Figure 1.26K) as a function of the excitation pulse duration. For fast-diffusing proteins like Halo-3xNLS and H2B-Halo, motion-blurring resulted in a large underestimate of the free diffusion constant, whereas the effect on slower proteins like CTCF and Sox2 was minor (Figure 1.26J). Regarding the total bound fraction, motion-blurring caused a $\sim$ 2 fold overestimate for rapidly diffusing Halo-3xNLS (Figure 1.26K), but had a minor effect on slower proteins like H2B, CTCF and Sox2. Similar results were obtained for both dyes for proteins with a significant bound fraction, but we note that JF_{549} appears to better capture the dynamics of proteins with a minimal bound fraction such as Halo-3xNLS (Figure 1.26J–K). Finally, we note that the extent of the bias due to motion-blurring will likely be very sensitive to the localization algorithm. Here, using the MTT-algorithm (Sergé et al. 2008), motion-blurring caused up to a 2-fold error in both the D_{FREE} and F_{BOUND} estimates.

   Taken together, these results suggest that Spot-On can reliably be used even for SPT data collected under constant illumination provided that protein diffusion is sufficiently slow and, moreover, provides a helpful guide for optimizing SPT imaging acquisitions (we include a full discussion of considerations for SPT acquisitions and a proposal for minimum reporting standards in SPT in Anders S. Hansen, Woringer, et al. 2018; Appendix 3 and 4).

1.3.7 Discussion

In summary, SPT is an increasingly popular technique and has been revealing important new biological insight. However, a clear consensus on how to perform and analyze SPT experiments is currently lacking. In particular, 2D SPT of fast-diffusing molecules inside 3D cells is subject to a number of inherent experimental (Figure 1.18A–B) and analysis (Figure 1.18C–D) biases, which can lead to inaccurate conclusions if not carefully corrected for.

Here, we introduce approaches for accounting for both experimental and analysis biases. Several methods are available for localization/tracking (Chenouard et al. 2014; Sergé et al. 2008) and for classification of individual trajectories (Monnier, Barry, et al. 2015; Persson et al. 2013). Spot-On now complements these tools by providing a bias-corrected, comprehensive open-source framework for inferring subpopulations and diffusion constants from pooled SPT data and makes this platform available through a convenient web-interface. This platform can easily be extended to other diffusion regimes (Metzler, Jeon, et al. 2014; an approach is presented in section IV.2) and models (Antony Lee et al. 2017) and, as 3D SPT methods mature, to 3D SPT data (section 1.2. Other extensions of Spot-On are presented in section IV.2 and in section IV.3. Spot-On is published as an open-source software[2].

[2]Computer code: Spot-On is fully open-source. The web-interface can be found at: https://SpotOn.berkeley.edu.

Moreover, spaSPT provides an acquisition protocol for tracking fast-diffusing molecules with minimal bias. We hope that these validated tools will help make SPT more accessible to the community and contribute positively to the emergence of 'gold-standard' acquisition and analysis procedures for SPT.

About one and a half year after the publication of Spot-On, several labs have across the world started to use Spot-On. Several metrics can be used to determine the popularity of the tool. First, Spot-On has been cited in 9 publications. Second, the official Spot-On server has seen about 3000 visits from all over the world (figure 1.27) since its launching, a number known to be an underestimate, since computers with privacy blockers installed are not counted.

Moreover, $\sim$ 1500 analyses were created on this server. This number is also known to be an underestimate for Spot-On usage, since Spot-On can also be used standalone using either a Python or Matlab interface.

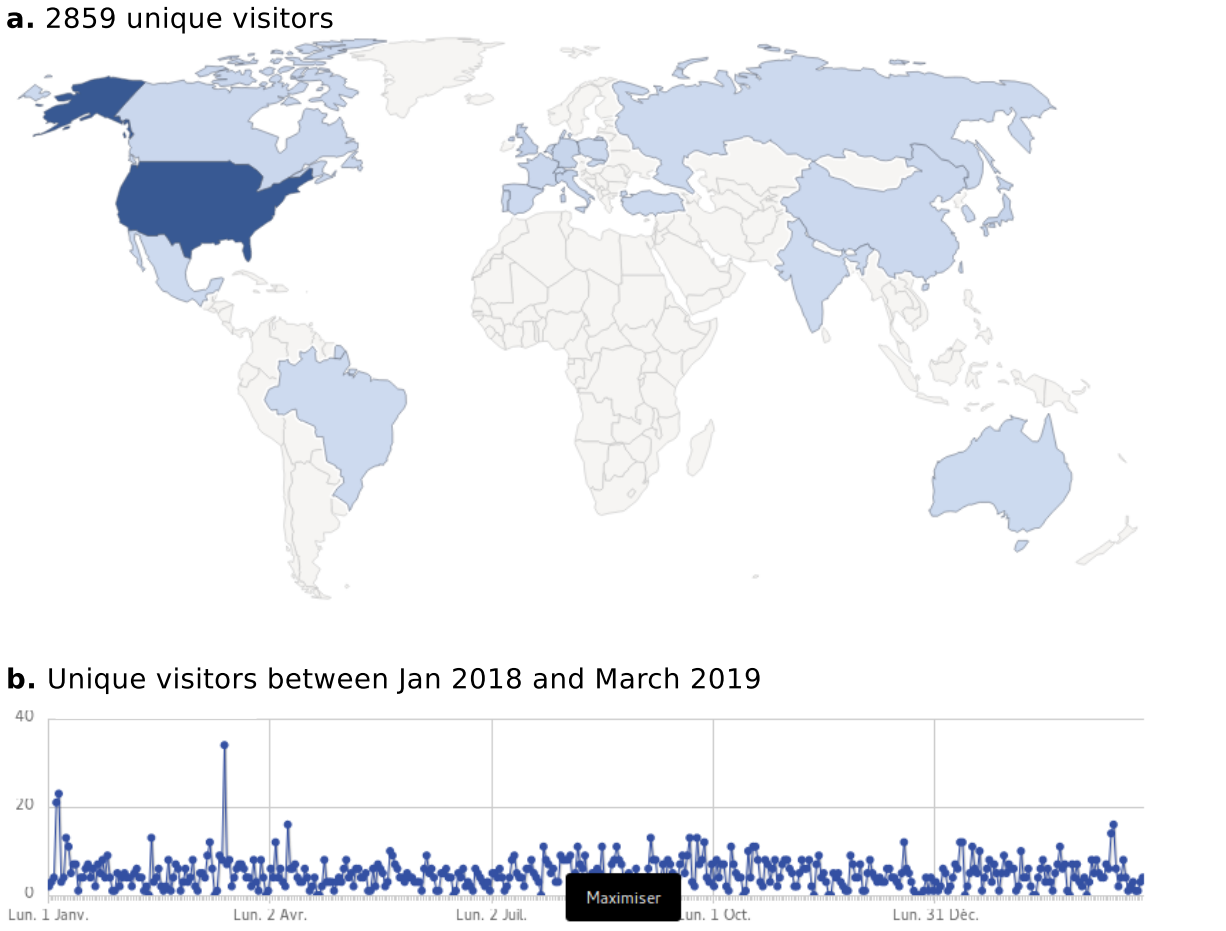


Figure 1.27: **Frequentation of the Spot-On website.** **(a)** By country. **(b)** Over time.

All raw code is available at GitLab: https://gitlab.com/tjian-darzacq-lab. The web-interface code can be found at https://gitlab.com/tjian-darzacq-lab/Spot-On; the Matlab command-line version of Spot-On can be found at: https://gitlab.com/tjian-darzacq-lab/spot-on-matlab; the Python command-line version of Spot-On can be found at https://gitlab.com/tjian-darzacq-lab/Spot-On-cli; the SPT simulation code (simSPT) can be found at: https://gitlab.com/tjian-darzacq-lab/simSPT; finally, the 'TrackMate to Spot-On connector' plugin, which adds an extra menu to TrackMate which allows one-click upload of datasets to Spot-On can be found at: https://gitlab.com/tjian-darzacq-lab/Spot-On-TrackMate.

Chapter 2

In vitro application: catalysis-enhanced diffusion

Introduction Spot-On (section 1.3) was developed with the goal of analyzing in-nucleus single-particle tracking data. It accounts for several biases occuring when a 3D medium is imaged in 2D. Furthermore, because Spot-On models general Brownian diffusion, it can also be used outside the context of a cell.

Indeed, Spot-On is expected to perform well in the presence of short trajectories, a setting where MSD is known (Michalet and Berglund 2012) to perform very poorly (see discussion in Appendix 1 from Anders S. Hansen, Woringer, et al. 2018). As such, we decided to apply the Spot-On to study the diffusion of enzymes freely diffusing in solution (*in vitro* diffusion), a setting in which diffusion is expected to be totally Brownian. This experiment served two goals:

- To provide a positive control to show that Spot-On actually models realistic jump length displacements, and that the deviations observed are due to deviation from Brownian motion inside the cell rather than our model not properly describing Brownian diffusion.
- To provide rare, additional data to the field of catalysis-enhanced diffusion, a field that studies the impact of enzyme catalysis to its diffusion coefficient.

This was a collaborative effort between Alan Shaw and Zhijie Chen (both affiliated with the Marqusee lab and the Bustamante lab, UC Berkeley), who were trying to perform SPT acquisitions. They prepared all the reagents (enzyme purification and characterization) and we discussed together the SPT experimental plan and I acquired the SPT images. Some of the sections of this chapter are extracted from a manuscript in preparation.

2.1 The catalysis-enhanced diffusion controversy

At the nanoscale, passive Brownian diffusion dominates the mobility of molecules. Molecular motors such as RNA polymerase can hydrolyze NTPs and generate force for directional movement along template DNA. Whether freely-diffusing enzymes can also harness chemical energy to generate additional mobility on top of Brownian motion is not well understood (Günther, Börsch, and Fischer 2018). Such a possibility seems to be supported by recent fluorescence correlation spectroscopy (FCS) measurements, which have shown that a number of non-motor enzymes including F1-ATPase (Börsch et al. 1998), urease (Muddana et al. 2010; Riedel et al. 2014; Jee et al. 2018), catalase (Riedel et al. 2014; Samudra Sengupta et al. 2013), ALP (Riedel et al. 2014), fructose bisphosphate aldolase (Illien et al. 2017), acetylcholinesterase (Jee et al. 2018) and hexokinase (Xi Zhao, Palacci, et al. 2018), enhance their diffusivities in the presence of substrates.

Accordingly, various mechanisms including fluctuations in pH (Muddana et al. 2010), global temperature increase of the solution (Golestanian 2015), force and charged product generation (Muddana et al. 2010), and enzyme chemotaxis toward substrates (Samudra Sengupta et al. 2013), have been

proposed to account for this phenomenon. Using FCS, the Bustamante and Marqusee labs previously uncovered a mechanistic link between the enhanced diffusion of catalase, urease and ALP, and the heat released in those exothermic reactions (Riedel et al. 2014). Within the framework of a stochastic theory, we proposed a "chemoacoustic effect" in which the heat released during catalytic turnover generates an asymmetric pressure wave that displaces the centre-of-mass of the enzyme, manifesting as catalysis enhanced enzyme diffusion. Arguing against this hypothesis, (Illien et al. 2017) further shown that aldolase, an enzyme that catalyzes a slow and endothermic reaction, exhibits enhanced diffusion in the presence of its substrate or competitive inhibitor. They proposed that the enhanced diffusion of aldolase is independent of the overall turnover rate of the reaction, but due to conformational fluctuations that alters the enzyme's hydrodynamic radius. More surprisingly, inert passive tracer has also been shown to diffuses faster in active enzyme solutions (Xi Zhao, Dey, et al. 2017), suggesting that the energy released during enzyme catalysis can be transferred to and harnessed by its environment. Recently, contrasting the aldolase diffusion enhancement observed by Illien et al., Zhang et al. studied the diffusion of aldolase with dynamic light scattering (DLS) and observed no diffusion enhancement in the presence of saturating concentration of substrates (Y. Zhang et al. 2018). DLS measures the D of molecules without the need of fluorescence, therefore its readout is independent of dye photophysics. These results are in direct conflict with one and another; and it's unknown what caused these discrepancies. To date, the vast no unified theory has been proposed to rationalize these experimental observations, which have all been made using FCS (Günther, Börsch, and Fischer 2018).

To date, the vast majority of publications describing enzyme diffusion enhancements, either experimental or theoretical, dependent on the validness of FCS measurements, which are sensitive to dye photophysics, and mounting evidences from non-FCS measurements and theory start to challenge the foundation of this field (Günther, Börsch, and Fischer 2018; Y. Zhang et al. 2018; Feng and Gilson 2019). In FCS, time traces of photons emitted while fluorescently labeled enzymes traverse a diffraction-limited confocal volume are recorded and analyzed with the autocorrelation function to derive the enzymes' ensemble-averaged translational diffusion coefficient (D, $\mu m^2/s$) (Krichevsky and Bonnet 2002). As such, factors other than fluorophore diffusion that cause fluorescence fluctuations within the confocal volume can interfere with D determination. Unstable fluorophore emission, fluorophore quenching, subunit dissociation of enzymes in diluted concentrations, and sticking of enzymes to surface of the coverslip can all contribute to erroneous D (Günther, Börsch, and Fischer 2018). Indeed, failure to rule out some of these factors have led to false claims in the diffusivity enhancement of F1-ATPase, which sticks to surface of the coverslip and undergoes substrate-induced dissociation of enzyme subunit (Günther, Börsch, and Fischer 2018; Börsch et al. 1998; Shah et al. 2013). Similar subunit dissociation has also been observed for aldolase (Woodfin 1967). Additionally, p-nitrophenyl phosphate (pNPP), a common substrate of ALP, has been shown to decrease fluorophore lifetime (Günther, Börsch, and Fischer 2018). These effects prompted us to cross-validate existing FCS observations with alternative techniques, ideally ones that allow the determination of D at the single molecule level and suffer less from dye photophysics.

SPT has emerged as a powerful approach to track the movement of individual molecules (Shen et al. 2017). By imaging fluorescent molecules at high speed and tracking the same molecule that appears in successive frames, trajectories of single molecules can be faithfully reconstructed. The jump length distribution from these trajectories can then be fit to theoretical models to derive D (Anders S. Hansen, Woringer, et al. 2018). Unlike FCS, this method is less sensitive to dye photophysics, as dark state or blinking fluorophores will not appear in successive images for reconstruction of diffusion trajectories.

ABEL trap is an advanced single-molecule technique that gives real-time position sensing via fluorescence detection and applies electric field feedback to counteract the molecule's Brownian motion, allowing $\sim$ 1-10 s continuous trapping—an observation time window three to four orders magnitude longer than that in FCS—of single molecules of various sizes (Quan Wang and W. E. Moerner 2011; Quan Wang and W E Moerner 2014; Quan Wang, Goldsmith, et al. 2012; Adam E. Cohen and W. E. Moerner 2005; A. E. Cohen and W. E. Moerner 2006). The voltage feedback is then used to reconstruct the molecule's diffusion trajectory and to derive D of single molecules (Q. Wang and W. E.

Moerner 2010). Apparent advantages of ABEL trap over FCS in measuring diffusion are that it 1) is independent of dye photophysics; 2) resolves sample heterogeneity such as different oligomeric states (Quan Wang, Serban, et al. 2018) and presence of free dyes; and 3) directly measures D of each individual molecules, what we term "single-molecule diffusometry" (Figure 2.1).

Here, we use SPT and ABEL trap to measure the D of single ALP molecules and show that catalysis, contrary to previous findings based on FCS, does not enhance their diffusivities at the single-molecule level. We uncovered surprising effect of pNPP-induced dye photophysics that mounted to artifact of 'apparent' diffusion enhancement of ALP in FCS. These new results urge a crucial revisit of previous FCS measurements on enzyme diffusion as well as various theoretical models, including those of our own, that relied on solely on FCS data. We further suggest use of control experiments, SPT, and ABEL trap as alternatives to substantiate other observations made using FCS.

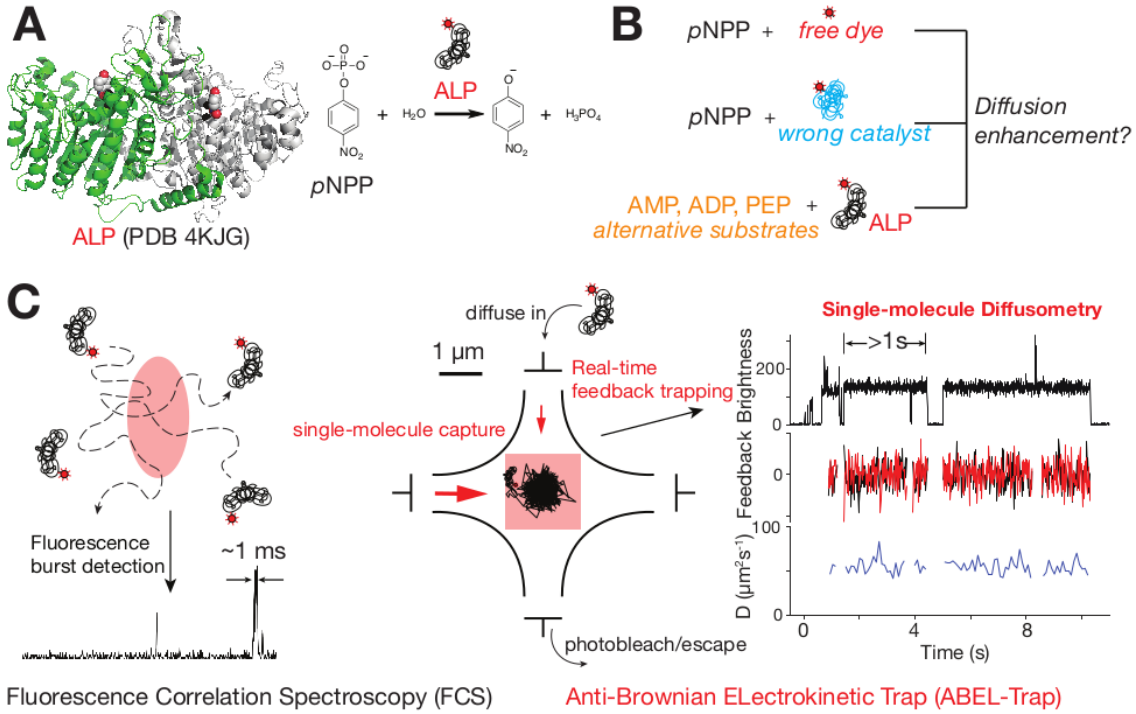


Figure 2.1: **Revisiting Catalysis Enhanced Diffusion of ALP** (A) Structure of ALP (left panel) and the reaction it catalyzes (right panel). The two protomers of ALP are colored in green and grey, respectively. pNPP and Zinc (black) are shown as spheres. During a reaction, ALP removes the phosphate group from pNPP and it has been proposed that catalysis enhances the diffusivity of ALP. (B) Control experiments including pNPP with free dye, pNPP with wrong catalyst, ALP reacting with alternative substrates are needed to examine the role of catalysis on ALP diffusivity. (C) Previous interpretation of enhanced enzyme diffusion are based on measurements from FCS, which detects the averaged fluorescence burst in confocal volume and is prone to photophysical artifacts. (D) ABEL trap uses electric feedback to counteract a molecule's Brownian motion and allows real time feedback trapping of a single biomolecule for seconds, an observation time window orders of magnitude longer than that of FCS. The feedback voltage is then used to reconstruct a molecule's diffusion trajectory, from which D of single molecules are derived (termed here 'single-molecule diffusometry').

2.2 Non-SPT experiments do not show enhanced diffusion

2.2.1 Catalysis is neither necessary not required for apparent diffusion enhancement of ALP in FCS

Among enzymes that have been reported to exhibit catalysis-enhanced diffusion, ALP shows the highest diffusion enhancement of up to 80% ($\Delta D/D_0$) with its substrate pNPP (Riedel et al. 2014) (Figure 2.1A). Such a big magnitude of catalysis enhanced diffusion deviates from the typical $\Delta D/D_0$ of 10-30% reported of other enzymes and is difficult to reconcile with theoretic predictions based on thermodynamic coupling between the swimming enzyme and its environment (Bai and Wolynes 2015). We therefore sought to investigate the enhanced diffusion of ALP by performing extra control experiments in FCS (Figure 2.1B) and use alternative techniques such as SPT and ABEL Trap to measure diffusion (Figure 2.1C).

We purified ALP from bovine intestinal mucosa to homogeneity through size-exclusion chromatography and fluorescently labeled the enzyme with JF_{646}-NHS. JF_{646} was chosen for its superior brightness and photostability (Jonathan B Grimm, Brian P English, Choi, et al. 2016), and that it has improved SPT and localization microscopy experiments (Basu et al. 2018). Using FCS, we recorded the raw photon intensity of ALP for 300 seconds, calculated the autocorrelation function $G(\tau)$ of every 10 seconds of the data, and fitted $G(\tau)$ with a simple model of diffusion of a single species in a dilute solution to derive D (Figure 2.2A-C). Averaging the diffusion constants over the 300-second time window gives a mean D of 45.6 $\pm$ 2.9 $\mu m^2/s$ (Figure 2.2C, data from 0 to 300s). Importantly, this analysis gives us an otherwise overlooked knowledge of the fluctuations of D over the experimental time window, which can be caused by interfering factors such as sticking of protein to the glass coverslip surface, large fluorescent spikes, evaporation of protein solution, and protein aggregation/dissociation. Addition of 2 mM pNPP to the same solution immediately quenched the raw fluorescence signal by $\sim$ 50 % (Figure 2.2A, right panel). The fluorescence signal recovered to 70 % of the initial intensity within 150 seconds and remained relatively stable thereafter. Such pNPP-induced fluorescence quenching effect also occurs on free dyes, as confirmed by quantifying fluorescence intensity of the free dyes in bulk using a fluorometer. Performing the same D analysis to the 300 seconds of data obtained post pNPP addition gives a mean D of 56.1 $\pm$ 4.7 $\mu m^2/s$ (Figure 2.2C, data after 300s), corresponding to a 23 % apparent diffusion enhancement. Notably, the fluctuation of difusion coefficients is greatly enhanced post pNPP addition (Figure 2.2C, red crosses). Consistent with the previous finding (Riedel et al. 2014), the apparent D enhancement in FCS is pNPP concentration dependent (Figure 2.2D, the first 4 groups with pNPP). Experiments with addition of buffer only did not give rise to any apparent D enhancement, demonstrating that the apparent D enhancement was not due to perturbation of the system during sample addition (Figure 2.2D, buffer group).

To test whether the apparent D enhancement of ALP in the presence of 2 mM pNPP originated from enzyme catalysis, we carried out several crucial control experiments. First, we measured the D of free JF_{646} dye before and after addition of 2 mM pNPP. As shown in Figure 2.2E, pNPP addition induced significant D fluctuations and 15% apparent D enhancement of free JF646 dyes (Figure 2.2E, free dye group), similar to those seen in experiments with ALP-JF_{646}. Second, we measured the D of Atto647N-labeled *Streptococcus gordonii* inorganic pyrophosphatase (Ilias and Young 2006) (sgPP-Atto647N) before and after addition of 2 mM pNPP. As pNPP is not a substrate of sgPP-Atto647N (i.e. wrong protein for catalysis), there should be no catalysis in this situation. However, we again observed similar D fluctuations and 17% apparent D enhancement (Figure 2.2E, sgPP group). Third, we measured D of ALP-JF_{646} in the presence of other ALP substrates including 2 mM AMP (Adenosine monophosphate), 4 mM ADP (Adenosine diphosphate) and 2 mM PEP (Phosphoenolpyruvate). ALP is fully active with these substrates but no apparent D enhancement is observed with any of them (Figure 2.2D, ADP, AMP, PEP groups). The low level of D fluctuations over time also suggest that these compounds did not interfere with D measurement in FCS. Unlike pNPP, these compounds do not quench the fluorescence of JF646 dye in bulk. Collectively, these data indicate that catalysis is neither necessary nor required for the apparent D enhancement of ALP in the presence of pNPP

(Figure 2.2D-E).

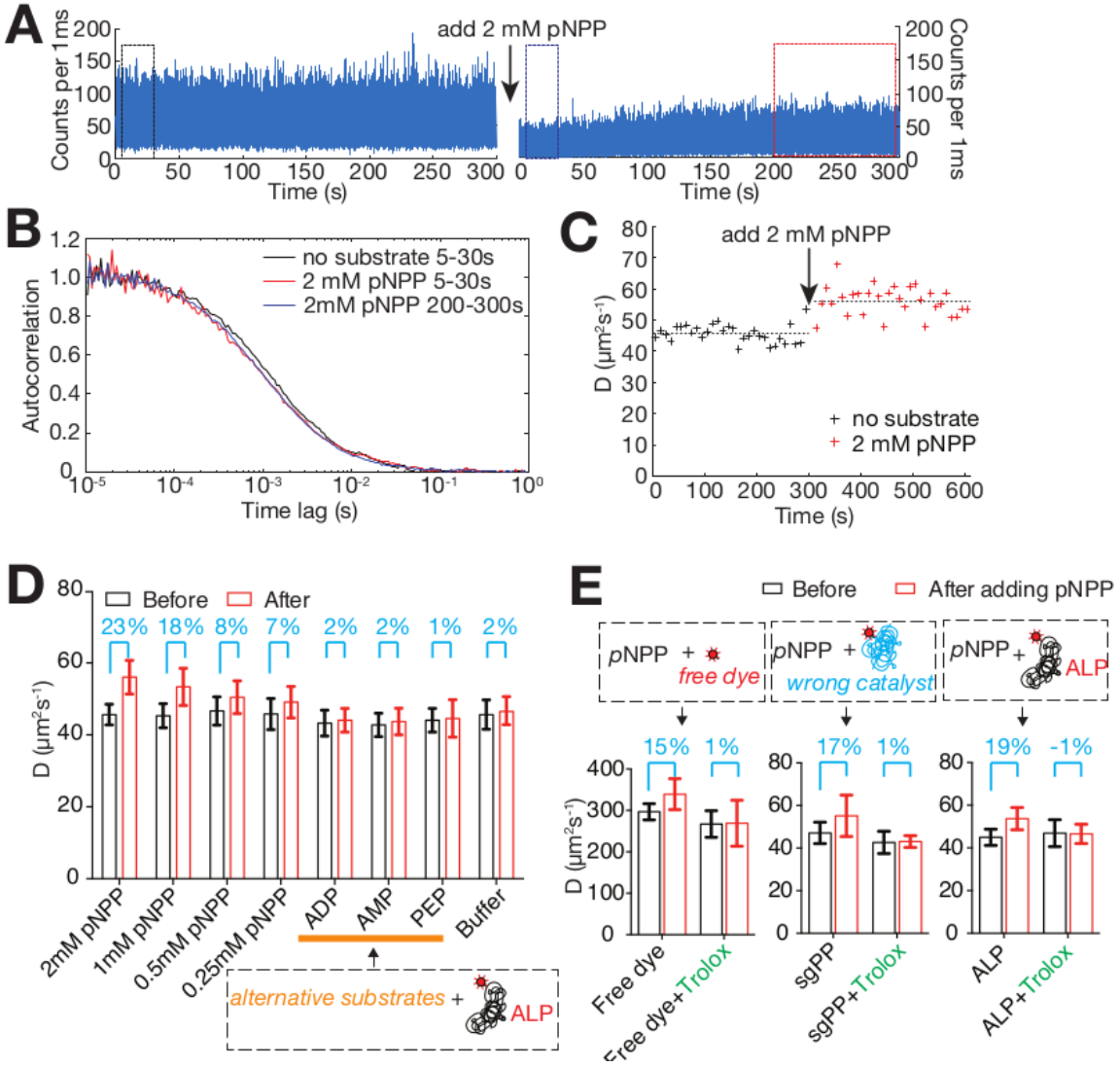


Figure 2.2: **Catalysis is Neither Necessary nor Required for Enhanced Diffusion of ALP in FCS** (A) FCS time trace of ALP-JF646 before (left) and after (right) adding 2 mM pNPP. (B) Normalized autocorrelation curves of data in black, blue and red rectangle boxes in (A). (C) Ds derived from fitting every 10s of data in (A). After adding 2 mM pNPP, the fluctuations of Ds become larger. (D) pNPP, but not other substrates causes apparent diffusion enhancement of ALP. Mean D of 300s of data before (black bars) and after (red bars) adding the indicated solutions. Error bars are standard deviations of 30 Ds. The percentage of mean D enhancement under each condition is labeled in cyan above the corresponding bars. The orange horizontal line indicates experiments with alternative substrates of ALP. (E) pNPP not only causes apparent diffusion enhancement of ALP, but also that of free dye and wrong catalyst. The apparent diffusion enhancement can be abrogated with a dye triplet state quencher, Trolox. The data is processed and represented similar to that of (D).

2.2.2 pNPP induces dye quenching and blinking

To directly measure the D of ALP at the single molecule level, we trapped single ALP-JF_{646} molecules in ABEL trap (Figure 2.3A). Most ALP molecules carry one or two dyes and the relative brightness of each molecule do not interfere with D estimation (Figure 2.3A). Extracting D from single molecules (Figure 2.3A, bottom panel) indicates that the enzyme exists as a single population in solution, with a narrow distribution of D (54.8 ± 0.2 $\mu m^2/s$, N = 216) (Figure 2.3A, right histogram). This value

matches well with D predicted from a hydrodynamic model (Basu et al. 2018) based on the enzyme's crystal structure (PDB 4KJG), suggesting that the enzyme behaves as a stable homodimer within our experimental time window ($<$ 2 hours) and enzyme concentration (20 pM).

Next, we repeated the trapping experiment under catalysis condition by adding 2 mM pNPP. To our surprise, we observed rapid and frequent blinking of the dye. Because ABEL trap relies on continuous photon detection to provide voltage feedback to counteract Brownian diffusion, the enzyme molecules therefore can no longer be stably trapped due to rapid loss of fluorescent signal. It is known that dye photobleaching and blinking can occur through oxidization or reduction of the dye triplet state into charge-separated states (Vogelsang et al. 2008). For example, a previous model on dye blinking and photobleaching suggests that both oxidizing and reducing agent of the dye triplet state are required for stable emission (Vogelsang et al. 2008). Accordingly, reducing and oxidizing systems (ROXS) are often used in single molecule fluorescence spectroscopy experiments to suppress these photophysical effects. Indeed, when we included a commonly used ROXS system (Trolox + PCA/PCD) in our trapping buffer, the pNPP-induced dye blinking effect was greatly ameliorated, allowing us to stably trap single ALP molecules in the present of pNPP (Figure 2.3D, left panel). Together, these results suggest that pNPP induces dye quenching and blinking, which likely acts through redox related pathways of the excited state of the dye.

2.2.3 ABEL trap reveals no catalysis enhanced diffusivity of ALP

Next, we trapped single ALP molecules in ABEL trap using the Trolox-PCA/PCD buffer (Figure 2.3C-D). Without substrate, ALP molecules showed a narrow distribution of D (53.3 $\pm$ 0.2 $\mu m^2/s$, N = 189) (Figure 2.3C, histogram on the right panel), similar to the above value determined in the buffer without Trolox-PCA/PCD (Figure 2.3A, right histogram of the top panel). In the presence of 2 mM pNPP, surprisingly, the distribution and mean of D (52.3 $\pm$ 0.2 $\mu m^2/s$, N = 198) (Figure 2.3D, histogram on the right panel) remained the same as those obtained without substrate. We confirmed that the enzyme remains active in all buffer conditions. Thus, the ABEL trap data does not agree with FCS results; and instead, suggest that there is no catalysis enhanced diffusivity of ALP at the single molecule level.

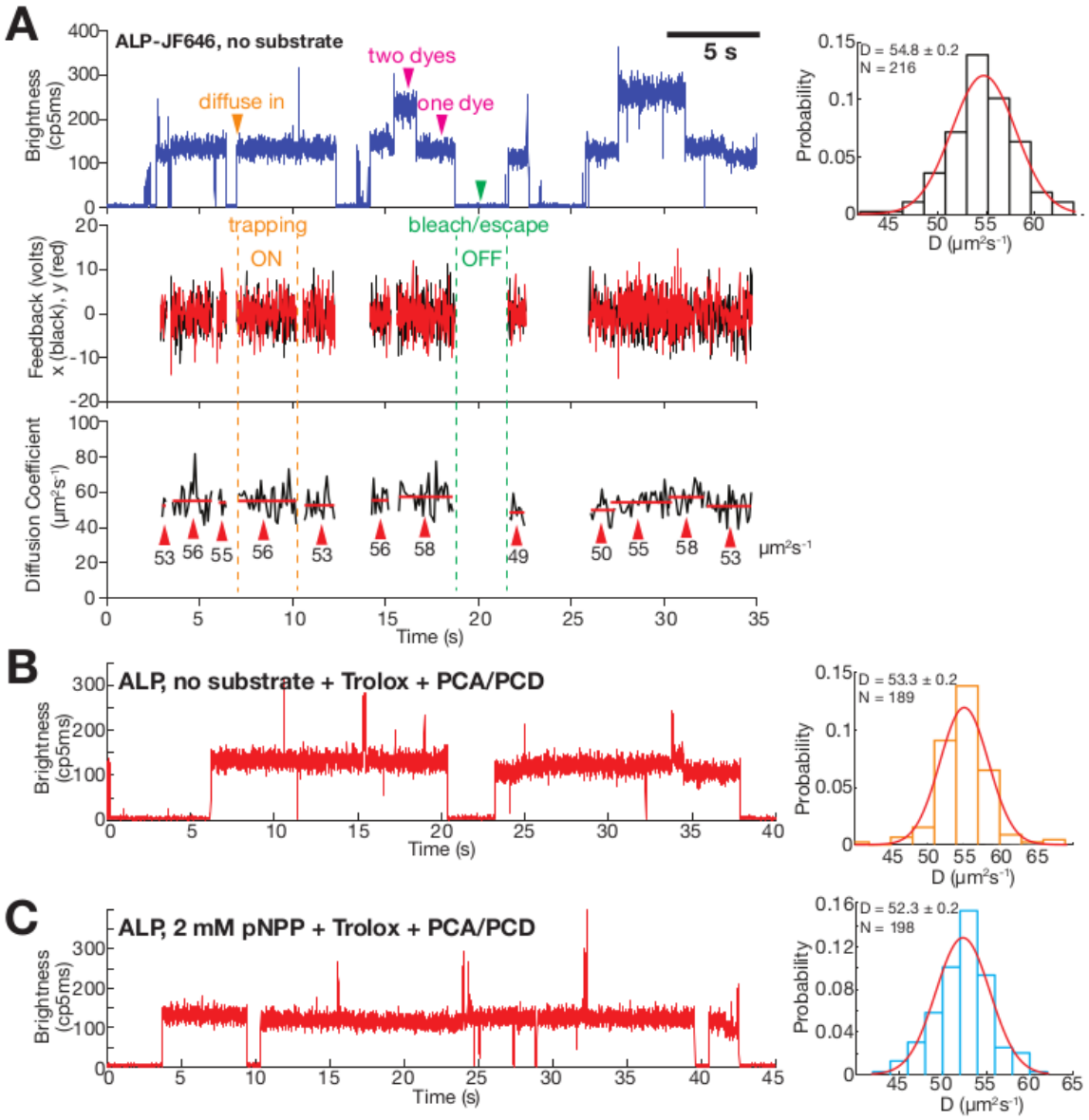


Figure 2.3: **ABEL trap Measurements Reveals no Catalysis Enhanced Diffusion of ALP** (A) A representative ABEL trap trace of ALP-JF646 with no substrate. Top left panel, intensity plot (photon counts per 5 ms) of the detected fluorescence signal. A single fluorophore corresponds to a signal of $\sim$ 130, whereas dual-fluorophore corresponds to $\sim$ 260, allowing the assignment of the number of dyes per molecule. Orange arrowhead (diffuse in) denotes the successful trapping of a single molecule. Magenta arrowheads denote the signal corresponding to one or two dyes. Green arrow denotes the loss of signal due to photobleaching or escape of the trapped molecule. Top right panel, D histogram of ALP-JF646 without substrate. Mean $\pm$ std of D and number of single molecules trapped (N) are displayed on the top left corner. Middle panel, the corresponding feedback voltage (x in black, y in red) applied in order to counteract Brownian motion and keep the molecule in trap. The feedback is on only when there is a molecule in the trap. The orange and green dashed lines denote timeframes when the trapping are on or off. Bottom panel, D of each trapped molecule calculated in real time (binned per 100 ms, black trace). The red lines indicate the mean D of each molecule, with the value written under the red lines and denoted with red arrowheads. The mean D of each molecule is used to derive the histogram on the top right panel. (B) A representative ABEL trap intensity trace (left) and D histogram (right) of ALP-JF646 with no substrate in the Trolox + PCA/PCD buffer. (C) A representative ABEL trap intensity trace (left) and D histogram (right) of ALP-JF646 with 2 mM pNPP in the Trolox + PCA/PCD buffer.

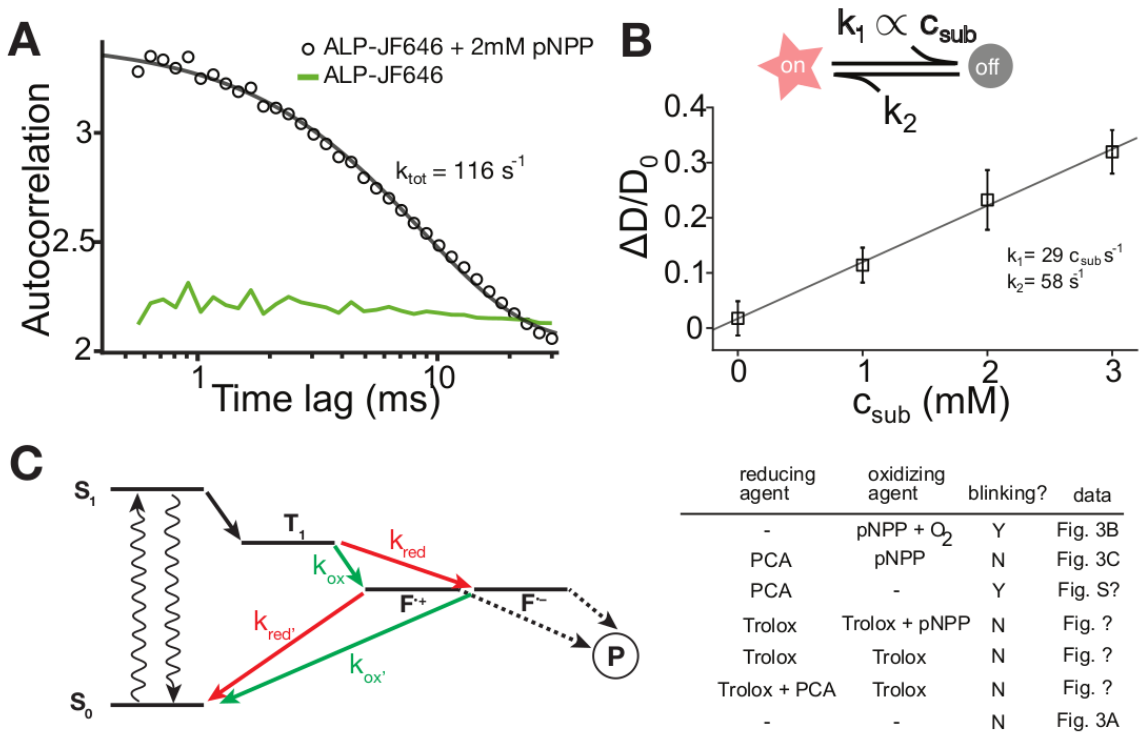
2.2.4 The apparent diffusion enhancement of ALP in FCS is caused by pNPP-induced photophysics of the dye

These new findings prompted us to reexamine previous FCS measurements. Because FCS measurements are sensitive to dye photophysics, we hypothesize that pNPP induced dye blinking and quenching, as observed in SPT and ABEL trap experiments, might play a role in the apparent D enhancement in FCS. To this end, we carried out FCS experiments in Trolox-PCA/PCD buffer, which alleviated pNPP induced dye blinking in ABEL trap (Figure 2.3C). Surprisingly, all the previously observed pNPP-induced apparent D enhancement of ALP-JF_{646}, free JF_{646} dye and sgPP-Atto647N were no longer evident (Figure 2.2E, + Trolox groups), lending support to the hypothesis that pNPP-induced dye photophysics contributes to the observed apparent D enhancement of ALP in FCS.

To test whether pNPP-induced fast dye blinking contributes to apparent D enhancement, we extracted the fluorescence intensity traces of ABEL trap data obtained in the absence and presence of 2 mM pNPP and fit them to autocorrelation. As shown in Figure 2.4A, pNPP induced millisecond time scale fluorophore blinking (k_{tot} = 116 s^{-1}) that could contribute significantly to fluorescence intensity fluctuations in confocal volume and D determination in FCS. We estimated the effect of blinking on the apparent diffusion enhancement using Monte-Carlo simulation. (Figure 2.4B).

To directly measure the effect of pNPP on dye photophysics, we conjugated JF_{646}-NHS dye to a 45-bp biotinylated dsDNA and immobilized the labeled DNA on the surface of a coverslip through streptavidin and biotinylated BSA (bovine serum albumin) interactions. We optimized dsDNA-JF_{646} concentrations until single fluorophores were sparsely distributed on the surface. In normal buffer conditions (1x PBS), we obtained stable emission of single dye molecules that bleaches in a single step. When 2 mM pNPP was added to the buffer, we observed complicated photophysical behaviors of the dye that varies from molecule to molecule but can be categorized into three distinct groups. The first group of molecules has a very short-lived fluorescence signal before it was quickly and irreversibly quenched. The second group of molecules shows unstable emission intertwined with blinking and stable emission. The third group of molecules shows fast blinking. These observations are consistent with the data from ABEL trap, confirming that pNPP indeed induced complicated photophysics of the dye. In addition, we repeated the surface experiment with Atto647 dye, which showed similar pNPP-induced fast dye blinking that was more consistent from molecule to molecule. These results indicate that pNPP may affect dye photophysics through a general mechanism.

To further explore the mechanism of pNPP's effect on dye photophysics (Figure 2.4C, left diagram), we conducted a series of trapping experiments with different ROXS conditions (Figure 2.4C, table on the right). Trolow has both oxidizing and reducing component, while PCA is a known reducing agent for the excited state of the fluorophore. In the presence of equimolar pNPP and PCA (2 mM each), the blinking was suppressed; while a concentration imbalance of pNPP (0.2 mM) and PCA (2 mM) caused rapid dye blinking. pNPP-induced dye blinking can also be suppressed by addition of Trolox (3 mM). Together, these experiments suggest that pNPP may acts as a dye excited state oxidizing agent, which alone quenches and causes blinking of fluorophores, but can work together with PCA and/or Trolox for stable dye emission (Figure 2.3C).



reducing agent	oxidizing agent	blinking?	data
-	pNPP + O_2	Y	Fig. 3B
PCA	pNPP	N	Fig. 3C
PCA	-	Y	Fig. S?
Trolox	Trolox + pNPP	N	Fig. ?
Trolox	Trolox	N	Fig. ?
Trolox + PCA	Trolox	N	Fig. ?
-	-	N	Fig. 3A

Figure 2.4: **pNPP-induced Dye Photophysics is Responsible for the Apparent D Enhancement of ALP in FCS** (A) Autocorrelation curves of ALP-JF646 intensity traces in ABEL trap without (green) and with 2 mM pNPP (black). The data are from trapping traces without Trolox + PCA/PCD. (B) Simulation of the pNPP-induced dye blinking kinetics into the diffusion of ALP-JF646 without substrate indicate a pNPP concentration dependent apparent D enhancement of ALP. (C) Left, a proposed model of pNPP-induced dye photophysics through a redox-mediated pathway of the dye triplet state. The dye triple state (T1) can be oxidized or reduced into charge separated states (F•+ and F•-) which is prone to photo bleaching (P, dark state). Rescuing the charge separated state with oxidizing or reducing agents returns the system to the ground state (S0) and suppress dye blinking and photobleaching. Right, ABEL trap data of ALP-JF646 with tuned redox condition of the trapping buffer. The data suggest that pNPP may act as a dye triplet state oxidizer and cause dye blinking. The photophysical effect of pNPP that can be suppressed with Trolox and/or PCA/PCD.

2.3 SPT reveals no catalysis enhanced diffusivity

2.3.1 Rationale

As an orthogonal approach to FCS and ABEL trap, we used SPT to characterize the diffusion enhancement of the enzymes in the presence of substrate. To do so, using the purified and labeled enzymes, we first characterized our microscope setup to estimate how fast we could observe diffusion. The, we performed acquisitions to measure the diffusion coefficient of the the enzyme with and without the substrate.

In order to obtain comparable results with FCS and to preserve the activity of the enzyme, we had to work in buffers of low viscosity. This required to optimize our SPT microscope to perform very high speed acquisitions. Indeed, whereas most transcripton factors observed in cells have a diffusion constant around 1-5 $\mu m^2/s$, enzymes in solution have a diffusion constant around 20-60 $\mu m^2/s$. To gain a factor of ten in the diffusion speed being captured, we took care of the following acquisition parameters:

- The passivation of the coverslips, in order to reduce out-of-focus fluorescence, and to be able to

image as close to the coverslip as possible (while keeping the coverslip out of the detection range, to avoid biases in the estimates of diffusion coefficients)

- The use of high laser power and short exposure time: we went down to 1.6 ms ($\sim$ 600 Hz) acquisition, in order to be able to track single molecules.

At that speed, most of the enzymes cross the axial detection range of the objective in about one frame, and most of the trajectories we obtain are composed of two points (median length: 1 frame). Thus, MSD-based techniques cannot be used to analyze such dataset. Conversely, Spot-On (section 1.3, because it builds a displacement histogram from two consecutive detections, can use the full information embedded in the data.

We assessed the diffusion of the following enzymes labeled with Atto-647N:

1. Catalase (an increase of 30% increase in D is expected)
2. Alkaline Phosphatase (ALP, increase of 80% in D expected)
3. Triose Phosphate Isomerase (TIM, $\sim$ 0%)

We present below the results obtained for these proteins. The results are also summarized in Table 2.1.

2.3.2 Diffusion of Alkaline Phosphatase (ALP)

To further confirm this surprising single molecule observation, we performed fast, in-solution SPT experiments and directly measured the diffusion coefficient of ALP with and without pNPP. In-solution SPT is an advanced single-molecule technique that allows tracking of individual fluorescently-labeled molecules over the course of several frames (Figure 2.5A) (Anders S. Hansen, Woringer, et al. 2018; Anders S. Hansen, Pustova, et al. 2017). Importantly, SPT is not affected by the photophysics of the dye, in particular by quenching as is the case with pNPP. Since the diffusion rate of freely-diffusing molecules in solution are above the tracking speed limit of our imaging setup, we assessed the diffusion of TIM-Atto647N dyes in buffers with different glycerol concentrations (10-25%) (Figure 2.6A). By tuning the viscosity of the buffer, we verified that the experimental D values of the dye in these buffers agree well with those predicted from theory (Figure 2.6B), confirming that this in-solution tracking strategy is sensitive enough for diffusion measurements. Moreover, we tested SPT of a fluorescently labeled protein, TIM-Atto647N (a protein with similar D to that of ALP), in a buffer with 10% glycerol. The data shows no differences in D with and without substrate (Figure 2.6C), confirming previous finding in FCS (Riedel et al. 2014). Next, we imaged ALP-Atto647N in a buffer with 10% glycerol, at $\sim$ 600 Hz (1.7 ms per frame) and used 1 ms stroboscopic illumination in order to minimize the motion blur of the protein during one exposure frame (Anders S. Hansen, Pustova, et al. 2017). The acquired traces (in the order of tens of thousands) were then analyzed using a population model (Anders S. Hansen, Woringer, et al. 2018) that expresses the distribution of jump length under a Brownian, free diffusion model (Figure 2.5B). Fitting both histograms of jump lengths to the model yielded a similar D with and without substrate (22.5 $\mu m^2/s$ with no substrate, and 23.3 $\mu m^2/s$ with 5 mM pNPP, Figure 2.5C). These two values differ by only 3.5%—within the experimental uncertainty, lending further support to the ABEL trap finding that catalysis does not enhance the diffusivity of ALP at the single molecule level. In the presence of pNPP, we detected approx 10x less particles (Figure 2.7A) and the spot intensity of detected particles were also dimmer (Figure 2.7B), consistent with the quenching effect of pNPP.

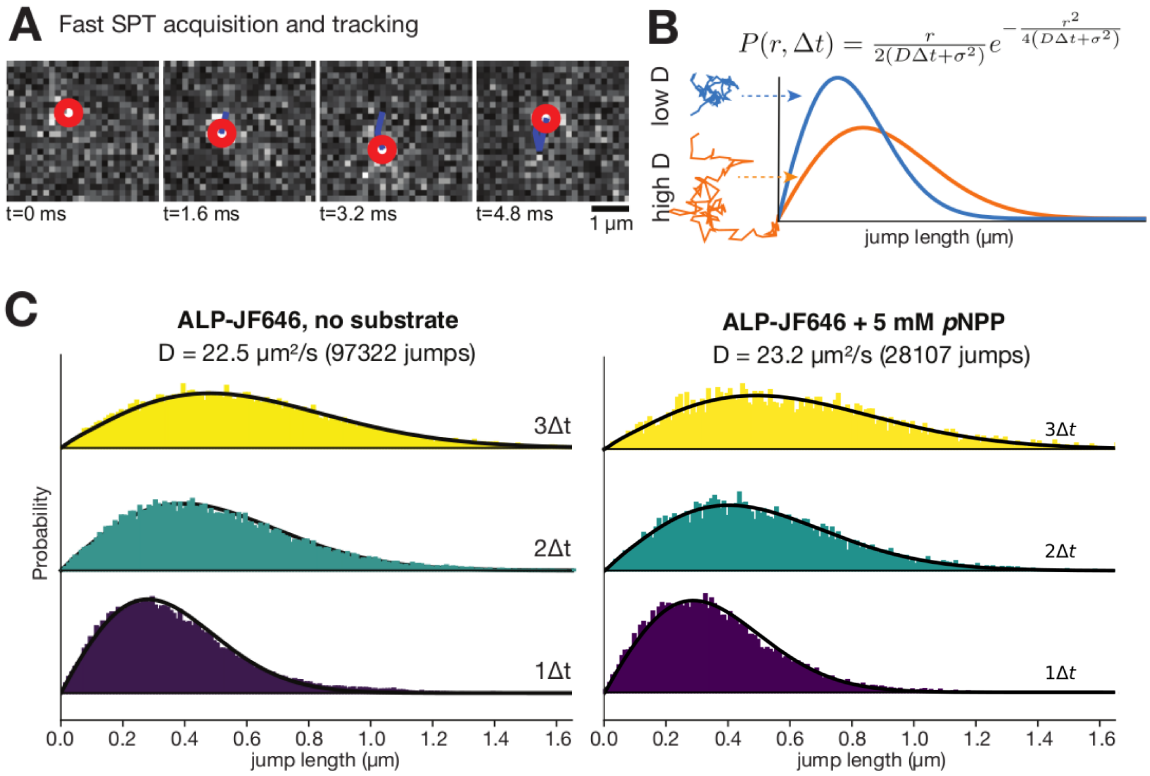


Figure 2.5: **SPT Reveals no Catalysis Enhanced Diffusion of ALP.** (A) Representative images of fast SPT acquisition and tracking of single molecules. The red circle denotes the center of the detected molecule in each frame. The blue line denotes the molecule's diffusion trajectory. (B) Modeling jump length distribution under free diffusion assumptions. Jump length distribution histograms of molecules with low and high D are plotted in blue and orange, respectively. (C) Jump length distribution histograms of ALP-Atto647N with no substrate (left) and with 5 mM pNPP (right). The yellow, green and purple histograms are distributions of jump lengths between 1Δ t (1.6 ms), 2Δ t and 3Δ t, respectively. The calculated D and total jumps are written on the top of the histograms.

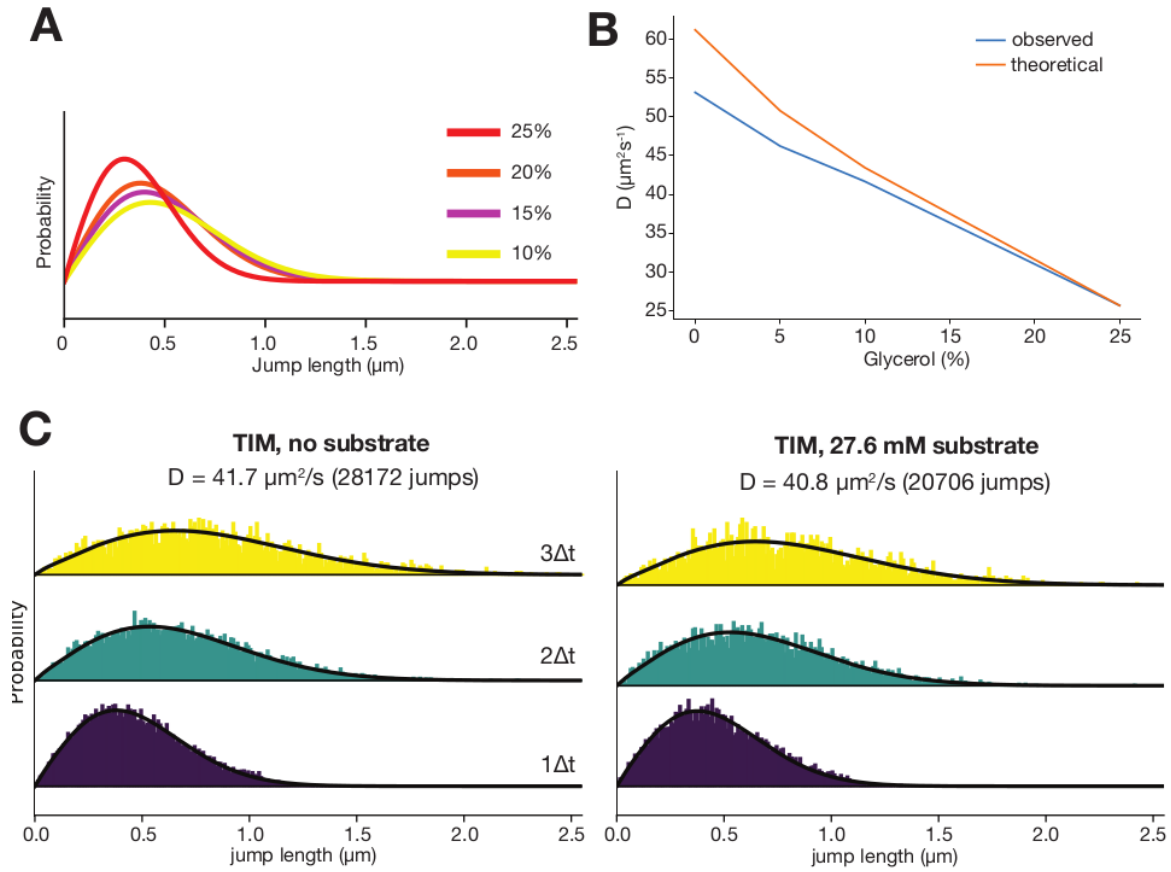


Figure 2.6: **Tuning Buffer Viscosity for in-solution SPT of an Enzyme** (A) Jump length distributions of TIM-Atto647N in buffers with different glycerol concentrations. (B) SPT calculated Ds of TIM-Atto647 match well with theoretically predicated values in buffers with different glycerol concentrations. (C) Jump length distribution histograms of TIM-Atto647N with no substrate (left) and with 27.6 mM D-glyceraldehyde 3 phosphate (right). The yellow, green and purple histograms are distributions of jump lengths between 1Δ t (1.6 ms), 2Δ t and 3Δ t, respectively. The calculated D and total jumps are written on the top of the histograms.

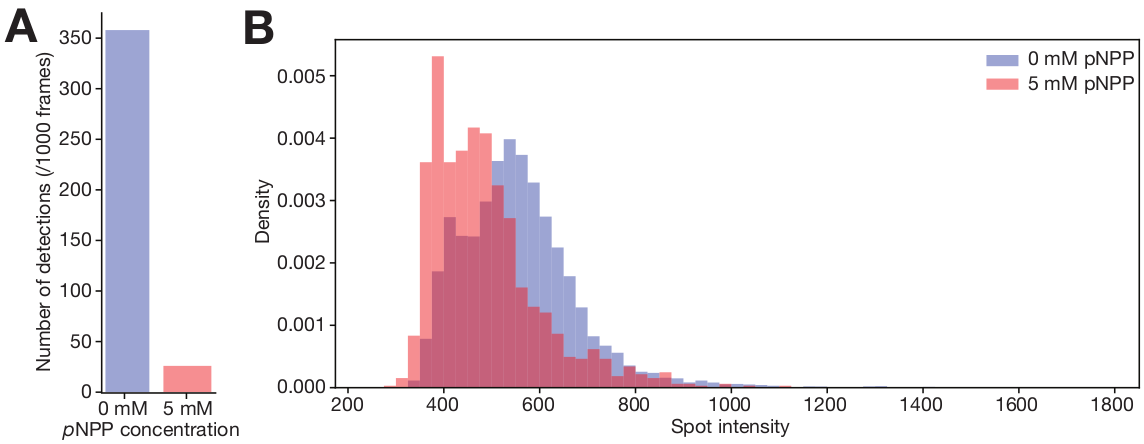


Figure 2.7: **pNPP quenches dye fluorescence in SPT** (A) Number of detected particles (per 1000 frames) of ALP-Atto647N without (pink) and with 5 mM pNPP (blue) in SPT measurements. Much fewer particles are detected in the presence of pNPP due to quenching or blinking of the fluorophore. (B) Histograms of detected spot intensites without (pink) and with 5 mM pNPP (blue). The spot intensities are lower in the presence of pNPP.

2.3.3 Diffusion of other enzymes

In addition to ALP, we also investigated the diffusive behaviour of other enzymes. We present results for triose phosphate isomerase (TIM), sgPP and aldolase in Table 2.1. Jump length displacements of sgPP are presented in Figure 2.8. For all the enzymes tested, SPT could not reveal any change in the diffusion coefficient.

Table 2.1: **Summary of SPT experiments with various enzymes and substrates. (∗)** Inconsistent results, **(∗∗)** Strong result ($\geq$ 2 replicates).

Enzyme	D (no substrate) ($\mu m^2/s$)	D (substrate) ($\mu m^2/s$)	[substrate] (mM)
TIM**	$\sim$ 38	$\sim$ 39	27.6
SGPP*	32	32	10
SGPP	32	33	10 (inhibitor)
Aldolase	33	33	10
bALP**	29	30	20 (PEP)
bALP**	29	28	5 (pNPP)
catalase	N/A	34	100

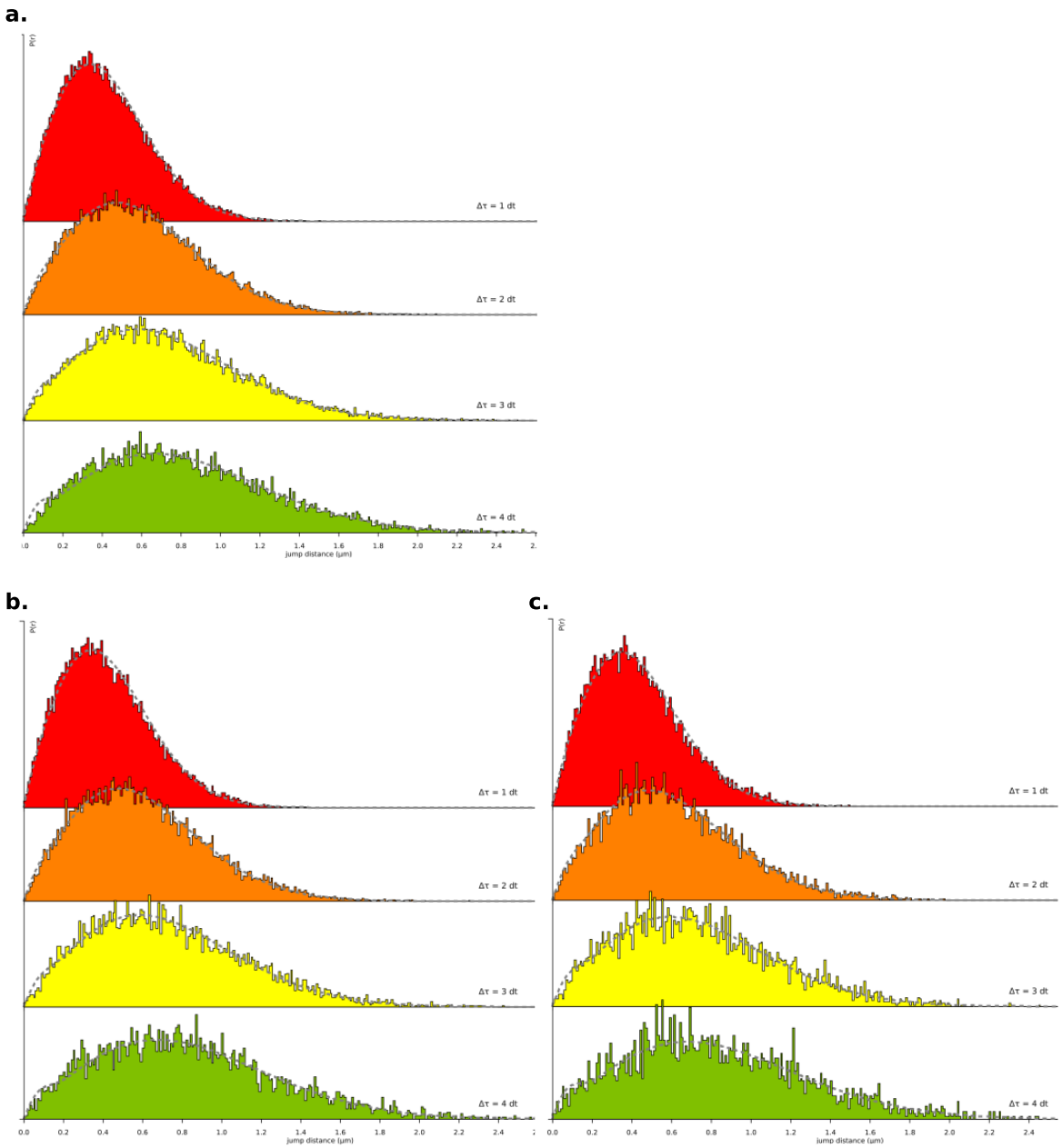


Figure 2.8: **Diffusion of sgPP.** Jump length displacement histograms at various time lags **(a)** In the absence of substrate. **(b)** In the presence of substrate. **(c)** In the presence of substrate and competitive inhibitor of sgPP. Quantified results are presented in Table 2.1.

2.4 Discussion

Propulsion of biomolecules at the nanoscale is no doubt a fascinating concept and has potential applications in the field of nanoscience and medicine. However, mounting experimental and theoretical evidences argue against the mechanism, scale and even the existence of such phenomenon (Günther, Börsch, and Fischer 2018; Y. Zhang et al. 2018; Feng and Gilson 2019). The vast majority of publications documenting enhanced enzyme diffusion upon catalysis were performed with FCS, which is prone to artifacts, such as free dye contamination, enzyme multimer dissociation and dye photophysics, that can result in false interpretations of enhanced diffusion, and often these artifacts are not accounted for in previous studies. Recently, Günther et al. characterized the impact of pNPP on the photophysics of Alexa 488, and concluded that pNPP reduces the fluorescent lifetime and quenches the dye in a

concentration dependent manner, resulting in a shorter correlation time measured in FCS (Günther, Börsch, and Fischer 2018). Zhang et al. utilized a non-fluorescent based technique, DLS, to monitor the diffusion of unlabeled aldolase upon catalysis, and showed that in contrast to previous FCS measurements, aldolase does not diffuse faster when catalytically active, and suggested that multimer dissociation may be the cause of the observed diffusion enhancement in FCS (Y. Zhang et al. 2018).

Here we show that, in agreement with Günther et. al., pNPP affects the photophysics of both JF_{646} and Atto647 dyes, which translates into a faster diffusion observed with FCS. Careful control of the buffer environment with ROXS can reduce the dye blinking significantly, allowing us to trap single ALP molecules with the ABEL trap in the presence of pNPP. Using both SPT and ABEL trap, we showed that at the single molecule level, the D of ALP does not increase in the presence of a saturating concentration of pNPP. By using several crucial control experiments, directly measuring dye photophysics, and simulation, we conclude that the apparent diffusion enhancement of ALP observed in FCS is primarily due to pNPP-induced dye photophysics. Although subunit dissociation was not observed in our ALP experiments, we note that ABEL trap is capable of resolving the diffusion of an enzyme with mixed multimeric states (Quan Wang, Serban, et al. 2018). Thus, we recommend the use of single molecule techniques such as SPT and ABEL trap that are not sensitive to dye photophysics to cross validate D measurements of fluorescent labeled biomolecules made in FCS.

Now that we have proved the validity and attractivity of Spot-On in such an *in vitro* setting, we decided to apply it to the study of a key transcription factor and oncogene: the c-Myc protein.

Chapter 3

In vivo application: sequence determinants of c-Myc dynamics

Introduction The nucleus of a mammalian cell is a highly complex organelle that governs the life of the cell through its ability to regulate transcription and metabolism, to integrate signalling responses and to control many division-related processes. In the past twenty years, our understanding of the nucleus has seen many revolutions, among them the rise of sequencing techniques (ChIP-seq and Hi-C) and imaging techniques (including FISH and live-cell imaging).

As detailed in section I.2, these tools helped to build a nuanced picture of the constitution and dynamics of the nucleus: it behaves as a complex aqueous medium filled with a high concentration of DNA, RNA, protein and other macromolecules. These molecules all exhibit some levels of interaction, from extremely weak to nearly-covalent bonds.

In this context, a key question in biology is to understand what are the biophysical parameters that determine the regulation of transcription. As explained in section I.3, the (potentially fractal) structure of nucleus and the surface characteristics of the TF both influence the dynamics of proteins.

Generally, the determinants of the dynamics of a TF (and furthermore its influence on transcription) are poorly understood. Even though it is thought that transcription efficiency is correlated with binding time to DNA (Clauß et al. 2017), our knowledge of the dynamics of TF is limited. Are protein-DNA interactions the main driver of a TF's dynamic? (in that case, one can expect significant changes in the dynamics when the DNA-binding domain of the protein is deleted). Contrary, and as suggested in (W. F. Lim et al. 2016), are protein-protein interactions the main drivers of the dynamics, and possibly of the choice of the DNA-binding location?

We decided to use the oncogene c-Myc as a model protein to address those questions. Briefly, c-Myc was chosen because it is a biomedically relevant protein (it is deregulated in most cancers, section 3.1.1 and is considered as a key therapeutic target, section 3.1.6), but also because preliminary experiments point to a unique behaviour in cells (section 3.1.7). In addition, it can also be used to ask basic biophysical questions about diffusion in the nucleus.

In this section, we decided to focus on understanding the sequence-determinants of the dynamics of the c-Myc protein. Concretely, we designed an experiment to evaluate how the dynamics of c-Myc changes when some of its protein domains are deleted. We motivate in more details this approach of the problem in section 3.2.1.

In order to make progress in this project, I received a huge amount of help from Lana Bosanac, Frank Liangqi Xie (genome editing attempts in NT2 cells), Claudia Cattoglio (biochemistry), Christian Weber (clonings and genomics), Gina Dailey (clonings), Claire Darzacq (biochemistry), Anders Hansen (SPT imaging) and Mickaël Lelek (microscope maintenance). Some preliminary experiments for this work were performed in 2014 and this section borrows some figures and text from my masters thesis (Woringer 2014).

3.1 Background, Myc as a...

3.1.1 Myc in cancer: discovery and epidemiology

1. **Myc as a prototypical oncogene.** Myc was initially discovered as an oncogene causing fulminant chicken tumor (Dang 2012), leading to the discovery of the v-Myc gene in the late 1970s. Soon after, orthologs genes were discovered and sequenced in several organisms, including mice and human (Bernard et al. 1983). A translocation involving the c-Myc gene in Burkitt lymphoma (a cancer affecting the white blood cells) lead to the discovery of this gene in humans in the early 1980s (Manolov and Manolova 1972; Dalla-Favera et al. 1982). More specifically, Myc is usually regarded as a proto-oncogene, since its transcription at normal expression levels does not usually lead to cancer.

2. **Myc epidemiology.** Since its discovery, Myc has been linked to many cancers. First, c-Myc causes Burkitt lymphoma through several chromosomal translocation. c-Myc is also mutated in many cancers (in 14% over all cancers; 20% of breast cancers, 30% of ovary cancers, almost 50% of some prostate cancers, etc.; Yinghua Chen et al. 2011; Kalkat et al. 2017), and several mutational hotspots have been identified in the protein (Nesbit, Tersak, and Prochownik 1999).

   Second, Myc expression has been shown to be elevated or deregulated in most (70%) human cancers (Beroukhim et al. 2010; Dang 2012), and is now associated with most cancers. Cancers in which Myc is deregulated are usually associated with "aggressive disease, metastatic potential, therapeutic resistance and poor patient outcomes" (Kalkat et al. 2017), especially when combined with other mutations or deregulations (Hartl 2016). As such, any drug that could repress Myc expression could globally improve the prognosis of the majority of cancer patients. In 1999, it was estimated that $\sim$ 70000 US cancer deaths were related to changes in c-Myc expression (Dang 1999).

3. **Various families of Myc.** Several related families have been discovered and include MYC (c-Myc), MYCL (L-Myc) and MYCN (N-Myc). Often, c-Myc and N-Myc can substitute for one another, in particular during murine development (Varlakhanova et al. 2010; Dang 2012). More precisely, the authors of (Varlakhanova et al. 2010) find that mice with c-Myc KO die in E10.5 and exhibit hematopoiesis and vascular defects (a link investigated in details in Kókai 2010). Furthermore, mice with a N-Myc KO die in E11.5 and display both lung and neuroectoderm defects. In general, the role of L-Myc is less clear, and L-Myc seems to be dispensable for mice embryonic development. Finally, when both c-Myc and N-Myc are deleted (dKO), the cells differentiate and lose pluripotency, suggesting a functional redundancy between c-Myc and N-Myc in mESC (K. N. Smith, A. M. Singh, and Dalton 2010; Fagnocchi and Zippo 2017).

4. **The Myc locus.** In this study, we focused on mice c-Myc. The c-Myc gene is located on chromosome 8 in human (gene identifier ENSG00000136997) and on chromosome 15 in mouse (ENSMUSG00000022346). It has three exons.

   c-Myc promoter. Transcription of the MYC gene can occur from several promoters named P0, P1, P2 and P3 (Bentley and Groudine 1986; Bardales et al. 2018; Carter, Jarquin-Pardo, and De Benedetti 1999). Transcripts starting from P1 and P2 seem to allow the production of the full-length protein. This first level of complexity at the promoter level is increased by the high number of enhancer sequences interacting with MYC.

   c-Myc enhancers. The regulation of c-Myc is thought to be extremely complex. First of all because the MYC promoter is surrounded and thought to be regulated by hundreds of enhancers (Enhancer Atlas lists 275 enhancers, T. Gao et al. 2016). Second, because some of the c-Myc enhancers are acting at extremely long distance, up to 1.7 Mb downstream, where most enhancers are located less than 100 kb from their target gene (von Paleske et al. 2014; Uslu et al. 2014). The complexity of this locus motivated the mapping of enhancers, first in a manual fashion (

Yochum 2011 for instance), and then through the development of specific enhancer screens using CRISPR interference (Fulco et al. 2016) and high-sensitivity 3C derivatives (Sumida et al. 2018).

3.1.2 The protein

1. **Protein domains.** The main c-Myc isoform is a 439 aa protein subdivided in several protein domains. As most transcription factors, c-Myc has a DNA binding domain and a trans-activation domain (Figure 3.1a).

   The *DNA binding domain* is a basic-region helix-loop-helix (bHLH) domain. This domain, characterized in (Fieber et al. 2001) is mostly unstructured when the protein is alone. However, the domain folds upon binding with its bHLH partner MAX, yielding what is considered as a fully functional DNA-binding domain that can be crystallized (Nair and Stephen K. Burley 2003 and Figure 3.1b). The rest of the protein is less ordered (see section 4), but has been divided in several Myc boxes (MB) based on functional characterizations. The bHLH domain is conserved across various branches, from human to *Drosophila*.

   The *Myc boxes* MBI, MBII and MBIII have been associated with transcriptional activation in deletion studies (Herbst, Hemann, et al. 2005; Cowling et al. 2006). MBII is a very conserved domain, from human to *Drosophila*. MBIV has been identified through conservation analysis. Despite being outside the DNA-binding domain, deletion of MBIV yields reduced DNA binding and reduced trans-activation (Cowling et al. 2006).

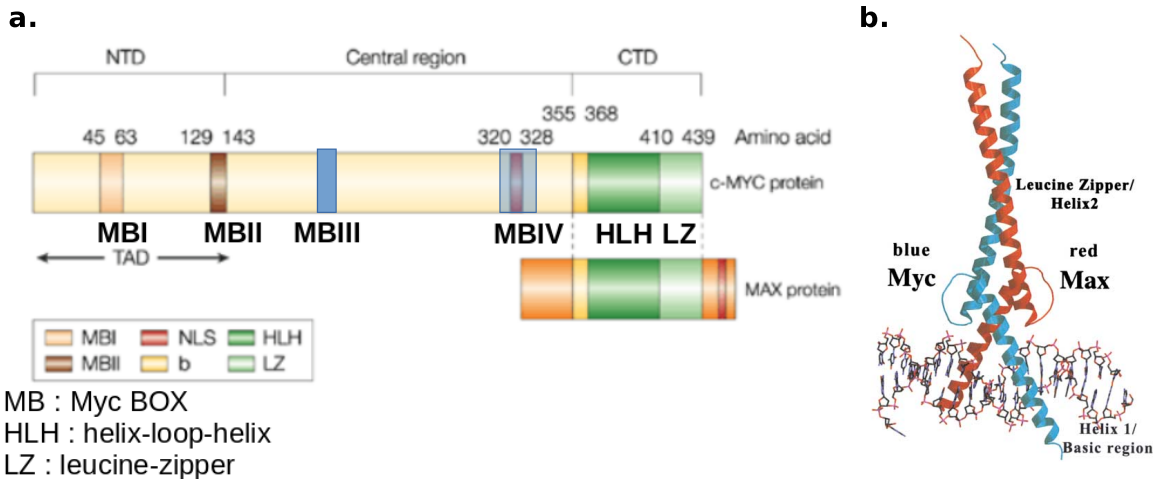


Figure 3.1: **Overview of the c-Myc protein. (a)** Domains identified in the c-Myc protein **(b)** Crystal structure of the CTD. source: adapted from (Pelengaris, Khan, and Evan 2002; Nair and Stephen K. Burley 2003)

2. **c-Myc isoforms.** The c-Myc protein can be found as two main distinct isoforms, termed Myc1 and Myc2. Myc1 starts at a non-canonical CUG codon and is 14 aa longer than Myc2, that starts from a regular AUG codon. Translation of Myc1 is initiated by a "scanning mechanism" whereas the translation of Myc2 is initiated by an Internal Ribosome Repositioning Element (IRPE) (Carter, Jarquin-Pardo, and De Benedetti 1999).

   Despite the minor difference in size (14 aa), the two isoforms are easily distinguishable by Western Blot for reasons that have not been totally elucidated (Myc1 runs at 70 kD whereas Myc2 runs at 64 kD). Furthermore, the ratio of Myc1/Myc2 concentrations has been shown to be regulated. Under normal cell culture growth conditions, the Myc1 isoform represents 10-15% of the total Myc pool. When cells are stressed by methionine deprivation or reach confluency, the Myc1 isoform gets preferentially translated (Stephen R. Hann, Sloan-Brown, and Gerald D. Spotts 1992). Finally, it has been proposed that Myc1 and Myc2 might be needed in precise proportions

to induce transcription, while the transcription of either Myc1 or Myc2 lead to the repression of the E-cadherin gene studied (Batsche and Cremisi 1999).

3. **Protein lifetime and stability.** Both the expression levels of the MYC mRNA and the c-Myc protein have been shown to be crucial for proper cell function. The stability of the c-Myc protein is tightly regulated through a complex mechanism of phosphorylation lock-in and its lifetime has been estimated around 20-30 minutes (GERALD D. Spotts and STEPHEN R. Hann 1990; Sears 2004; J Liu and Levens 2006).

   Concretely, c-Myc stability is controlled by the presence of post-translational modifications (PTMs, some of them represented in Figure 3.2) that target it to the proteasome for degradation. The two phosphorylations involved are Threonine 58 (Thr58) and Serine 62 (Ser62). When Thr58 is phosphorylated (by GSK3β), c-Myc is destabilized. This is only possible if Ser62 is already phosphorylated. Conversely, when Ser62 is phosphorylated (by ERK), c-Myc is stabilized. These two PTMs are located in a mutation hotspot present in cancer.

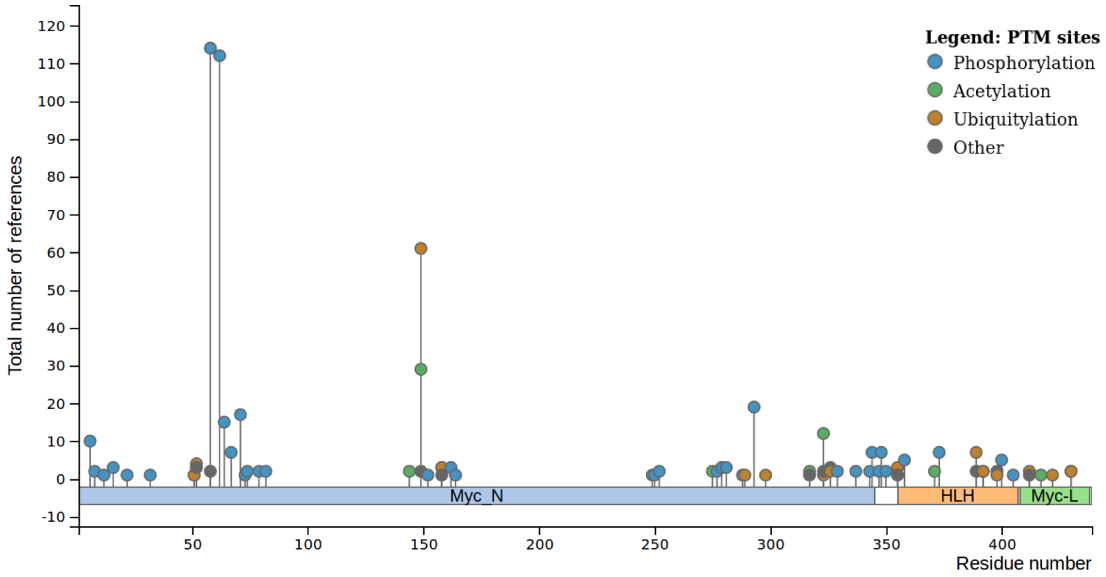


Figure 3.2: **Location of main c-Myc post-translational modifications (PTM)**. Source.

From a series of experiments, one can infer the following timing of events (summarized in Figure 3.3: (1) Ser62 is phosphorylated by ERK, (2) Thr58 is phosphorylated by GSK3β, (3) c-Myc proline 63 is isomerized by the prolyl-isomerase Pin1, (4) Ser62 is dephosphorylated by phosphatase PP2A, (5) c-Myc is ubiquitylated by Fbw7 and degraded by the proteasome (Yeh et al. 2004; Sears 2004; Garrison and Rossi 2010; X.-X. Sun et al. 2018).

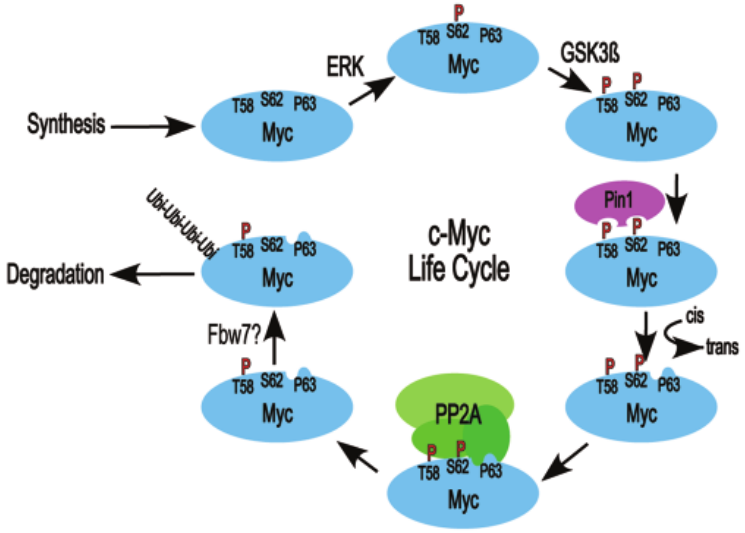


Figure 3.3: A proposed mechanism to regulate c-Myc stability and degradation. source: Sears 2004

4. **c-Myc structural dynamics.**

   Similar to most transcription factors (section I.3), Myc is entirely disordered in solution. When bound to its obligatory partner MAX, the DNA-binding domain, bHLH, of c-Myc folds upon binding, but the rest of the protein remains disordered. Both experimental and computational approaches have been used to estimate the disorder of the c-Myc protein, and a consensus is presented in Figure 3.4. Furthermore, following a classification established by Das, Ruff and Pappu (Rahul K Das, Ruff, and Rohit V Pappu 2015) based on the amino-acid sequence of the protein, c-Myc belongs to the category of "Janus proteins", intermediate between well-folded and disordered proteins, a classification corroborated by experimental studies (Figure 3.5).

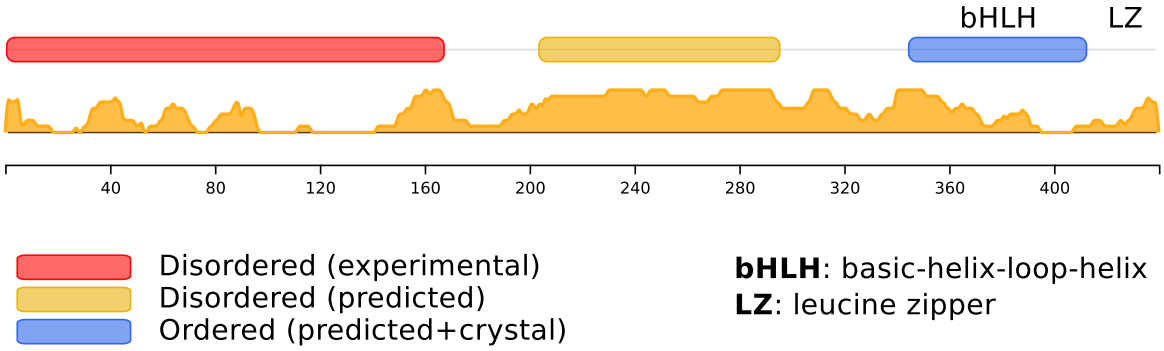


Figure 3.4: Disorder score across the c-Myc sequence, as predicted by a consensus method. source: `http://mobidb.bio.unipd.it/`

Some of the domains of c-Myc have been studied independently using various experimental or computational studies.

The bHLH domain of c-Myc was shown to have a low α-helical propensity (it tends to be disordered) and it very quickly dimerizes with MAX (Fieber et al. 2001). Moreover, a series of related bHLH domain were shown to be disordered when not bound to other proteins/DNA by both circular dichroism experiments and molecular dynamics simulations (Rahul K. Das, S. L. Crick, and Rohit V. Pappu 2012).

A region encompassing MBI (aa 1-88) was studied by nuclear magnetic resonance (NMR) chemical shift analysis, relaxation measurements and Nuclear Overhauser effect (NOE) analysis (An-

dresen et al. 2012) found that the region was mostly unstructured but seemed to interconvert with sub-regions of transient secondary structure. These sub-regions tend to be more conserved than the rest of the protein and are known to be involved in protein-protein interactions. Finally, based on these results, the authors also predicted that deletion of an entire domain might have adverse consequences, by bringing into contact two domains that might then adopt previously unexplored conformations.

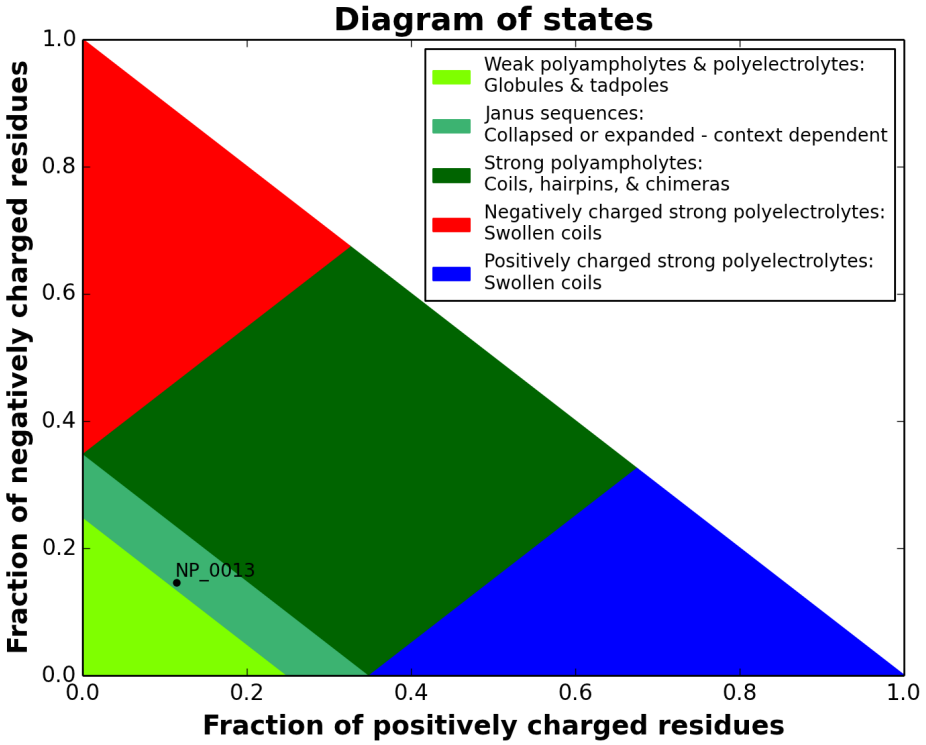


Figure 3.5: Location of c-Myc (NP_{0013}) within the Das & Pappu diagram, representing its propensity to be disordered. c-Myc is located in the region of "Janus proteins".

Finally, as many transcription factors and nuclear proteins, results suggest that purified c-Myc may undergo phase separation *in vitro* (Boija et al. 2018), a process that might be highly relevant for transcription regulation (section 3.2.3).

3.1.3 Oncogene or General Transcription Factor (GTF)?

The role of c-Myc as a transcription factor still remains elusive, despite several decades of intense investigation. Indeed, several questions remain open, including: the nature of the main c-Myc-interacting partners (section 2), the identity of c-Myc target genes remains controversial (section 3.1.4) and the mechanism by which c-Myc binds to specific binding sites (section 1).

1. **c-Myc binding site.** Soon after the sequencing of the MYC gene, *in vitro* experiments characterized the DNA affinity of the protein (Halazonetis and Kandil 1991) and found that c-Myc has a canonical binding motif `CACGTG`. The authors also found that c-Myc has weaker affinity for non-canonical, degenerate binding sequences: `CANNTG`. The genomic loci bearing this sequence were termed E-boxes. Sequences flanking this motif were shown to influence c-Myc *in vitro* affinity.

   This degenerate motif makes E-boxes an extremely abundant sequence in the genome (the canonical motif is expected to occur once every 4 kb and the degenerate motif once every 256 bp). As

it is common with transcription factors, ChIP-seq experiments quickly demonstrated that the knowledge of the transcription factor binding motif is not sufficient to predict c-Myc binding on the genome (Guccione et al. 2006), and that the specificity of c-Myc binding had to be examined along two additional axes: (1) the existence of partners that might restrict c-Myc binding (section sec:myc-partners) and (2) the chromatin accessibility.

2. **c-Myc interacting partners.** In order to understand how c-Myc can act as an oncogene, many of its protein interacting partners were progressively identified. One of the first partners identified was MAX (Amati et al. 1992). MAX is a bHLH protein that can either oligodimerize (MAX-MAX dimer) or form several heterodimers (cMyc-MAX, or MAD-MAX; Zada et al. 2006). In order to bind DNA, c-Myc has to form an obligatory dimer with MAX. Thus, the abundant protein MAX acts as a regulator of c-Myc: when MAX is sequestered in its interaction with proteins from the MAD family, it does not dimerize with c-Myc and the protein cannot transactivate its target genes.

   Beyond this obligatory interaction, c-Myc has been shown to interact with many other proteins or complexes. It interacts with histone acetyltransferases and deacetylases like TRRAP and GCN5 and has thus been suggested to play a role in keeping chromatin in an "open environment" (Eberhardy, D'Cunha, and Farnham 2000; McMahon, Wood, and Cole 2000; Frank et al. 2001; Kurland and W. P. Tansey 2008).

   Moreover, additional mass spectrometry-based screens identified or confirmed series of more than a hundred partners (Dingar et al. 2015). Some of them were further investigated, such as Mediator subunits, the SWI/SNF chromatin remodeling complex (Stojanova et al. 2016), the methyltransferase complex SET/MLL (Thomas et al. 2015), its link with histone modifiers (Lüscher and Vervoorts 2012) or many other core components of the transcription and replication machinery (including Aurora A, pTEF-b,RNA PolII, etc; Büchel et al. 2017; Myant et al. 2015, some of them reviewed in Tu et al. 2015).

3. **Towards a mechanism.** The extremely high number of c-Myc partners drew a very complex picture of its role as an oncogene, and many conflicting hypothesis co-exist to explain its mode of action.

   (a) *Myc as a general transcription factor.* A first series of hypotheses stemmed from ChIP-seq and pre-ChIP-seq experiments (Guccione et al. 2006; Zeller, XiaoDong Zhao, et al. 2006; Nie et al. 2012; Lin et al. 2012), found that c-Myc could be found bound to the promoters of virtually all active genes, and that the intensity of the c-Myc peak correlated with the expression level of the gene. These experiments thus pointed to the fact that c-Myc could almost be considered as a general transcription factor, very close to the core transcriptional machinery. These results were in agreement with experiments that found that c-Myc interacted with the positive transcription elongation factor (pTEF-b), a key complex involved in the regulation of of transcriptional RNA PolII pausing (Rahl et al. 2010). c-Myc was proposed as a factor promoting pause release, and thus favorizing transcription.

   These experiments resonated with the previous hypothesis that linked the interaction of c-Myc with histone acetyltransferases, leading to the idea that c-Myc might act as a global genome organizer that keeps chromatin in an open environment by recruiting histone modifiers to promoters (Niwa 2007; Knoepfler et al. 2006; Knoepfler 2008; Cotterman et al. 2008; Kieffer-Kwon et al. 2017).

   (b) *The metabolism hypothesis.* Another series of hypotheses emerged from the initial identification of c-Myc as a key metabolic regulator. Indeed, c-Myc has been implicated in the regulation of genes involved in key metabolic reactions, including the synthesis of purine and pyrimidine nucleotides and glucose metabolism (Dejure and Eilers 2017). As such, Dejure and Eilers developed the concept of Myc-induced metabolic reprogramming: in order for a

tumor to develop (tumors tend to rely more on anaerobic reactions than non-tumor cells), several Myc target genes involved in nucleotide synthesis and glucose metabolism need to be over-expressed.

3.1.4 The target gene controversy

Results presented in section (section 3.1.3) pointed to the fact that the role of c-Myc in several contexts could be better understood by dissecting its target genes. Indeed, the quest of c-Myc target genes spans more than two decades, and it is only this year that clear answers seem to emerge.

1. **Initial experiments.** The first c-Myc target genes were determined by trial and error with scientists making educated guesses about Myc potential target genes. As such, one of the first target gene identifies was the ornithine decarboxylase (Bello-Fernandez, Packham, and Cleveland 1993) gene. Other target genes were identified similarly (Dang 1999), and most of the efforts were focused on metabolism-related genes. A comprehensive, genome-wide picture was still lacking.

2. **First genome-wide screens.** The first genome-wide screens were performed using differential expression systems, in which the cell transcriptome was assayed with and without the c-Myc protein, or by using pre-ChIP-seq techniques (Menssen and Hermeking 2002; Zeller, Jegga, et al. 2003; Fernandez et al. 2003; Zeller, XiaoDong Zhao, et al. 2006; Seitz et al. 2011, reviewed in Dang et al. 2006). These studies confirmed that *in vivo*, c-Myc mostly binds to E-boxes. They lead to the identification of thousands of potential binding sites and thousands of potentially regulated genes without any clear unifying mechanism. In the late 2000s, the mechanism of action of c-Myc remained elusive.

3. **c-Myc as a genome-wide amplifiers.** In 2012, however, two back-to-back papers tried to provide the missing unifying framework (Nie et al. 2012; Lin et al. 2012). First, these papers confirmed that c-Myc binds at virtually all active genes. Furthermore, by performing RNA-seq with careful spike-in controls, the authors found that when c-Myc is over-expressed, the global RNA levels are increased: all expressed genes become over-expressed, leading to what has been named as a "global genome amplification".

   Further ChIP-seq-based studies found that when over-expressed, c-Myc tended to occupy previously unoccupied binding sites, termed "lower affinity binding-sites", and thus to activate new genes (Lorenzin et al. 2016), a vision that nicely complemented the two 2012 papers.

   These studies concluded that the question about c-Myc target genes had been phrased wrongly, since c-Myc seemed to regulate all genes. Furthermore, these studies provided an attractive explanation for the role of c-Myc in cancer: when c-Myc is over-expressed, all genes tend to be over-expressed (global genome amplification), leading to an expected cancer-like phenotype.

4. **From indirect to direct targets.** These unifying results were, however, rapidly challenged by several studies, mostly originating from the Bruno Amati group. Indeed, most of the previous studies could not distinguish between a direct and an indirect effect on target genes. Is the global genome amplification a direct consequence of c-Myc over-expression or a consequence of the activation of one of c-Myc target genes?

   First, a ChIP-seq-based study (Sabò, Kress, et al. 2014) identified classes of genes that responded differently to c-Myc over-expression in B cells. This started to challenge the "global amplifier vision". This study highlighted complex feedback mechanisms that might explain mRNA amplification. Second, using a Myc-inducible system, classes of repressed genes were identified when c-Myc was induced (Walz et al. 2014), painting a more complex picture, in which it is actually necessary to finely characterize c-Myc targets to understand its role as a transcriptional regulator. Along this line of evidence, the c-Myc-Max dimer was associated with a specific interaction with Polycomb repressive complexes (Krepelova et al. 2014; Benetatos, Vartholomatos, and Hatzimichael 2014).

Third, this vision was refined by a reanalysis that investigated which step of transcription was regulated (de Pretis et al. 2017), pointing to PolII loading as the c-Myc regulated step, a result that contrasts with previous papers hinting at pause release (Rahl et al. 2010).

Despite being more and more precise, these studies were still providing rather indirect evidence that could not distinguish direct from indirect target genes. The most recent studies to date used a quickly-inducible system that reacts in less than one hour, allowing to distinguish between direct effects of c-Myc induction and the subsequent activation cascade (Muhar et al. 2018; Sabò and Amati 2018). This study confirmed that c-Myc directly activates a distinct set of genes, and that many indirect, cascade effects occur later on.

3.1.5 c-Myc in pluripotency & development

1. **c-Myc and reprogramming.** In addition to the long-lasting interest in c-Myc due to its central position in cancer biology, c-Myc has experienced a renewed interest in 2006 when the Yamanaka group first generated induced pluripotent stem cells (iPSCs) by over-expressing a cocktail of four pluripotency-related factors: Oct3/4, Sox2, c-Myc, and Klf4 (Takahashi and Yamanaka 2006). This pioneer work demonstrated that only four variables (the concentration of four factors) was sufficient to de-differentiate fibroblasts in mES-like cells, and suggested that these factors could be master regulators in pluripotency and development. Reprogramming was later shown in human cells, and finally cruelly demonstrated inside live mice (Abad et al. 2013) This finding also made sense in light of c-Myc's interaction with chromatin modifiers an its alleged role in maintaining an "open chromatin environment".

   Further studies, however, found that despite the fact that reprogramming efficiency was improved in the presence of c-Myc, it was also possible to de-differentiate cells in the absence of c-Myc (Yeo and Ng 2013).

2. **c-Myc and pioneer factors.** Since c-Myc was a facilitating factor to induce iPSCs, questions arose about the mechanism of reprogramming: how could initially silent (and nucleosomized) sequences be reactivated?

   An explanatory model was proposed by the Zaret lab (Zaret et al. 2008). To be able to activate a previously nucleosomal locus, specific transcription factors need to be able to bind nucleosomes. These factors are termed pioneer factors. Despite its role in reprogramming, and unlike Oct4 and Sox2, c-Myc did not show any pioneer activity (Sekiya et al. 2009; Soufi et al. 2015).

3. **c-Myc during mice development.** We highlight here two aspects of the interaction between c-Myc and development, as an example of the amplitude of the cross-talks between c-Myc and developmental processes.

   c-Myc in the early embryo. Similar to Nanog (Filipczyk et al. 2015; Hastreiter and Schroeder 2016), c-Myc has been shown to display heterogeneous levels across cells at the same time, both in mESC and in the early embryo (Clavería et al. 2013; Díaz-Díaz et al. 2017). The reason of such natural heterogeneity is still unclear, but has been shown to be linked to cell survival, cell-to-cell competition and regulation of necrosis, all mechanisms that have been shown to be crucial in the early embryo.

   c-Myc during hematopoiesis. In addition to being involved redundantly with N-Myc in early embryo development, the role of c-Myc in hematopoiesis and angiogenesis has been investigated. c-Myc is necessary for the maintenance of bone marrow hematopoietic stem cells (Wilson 2004). Conversely, a conditional over-expression of c-Myc in mice leads to severe defects in the vascular system, including a reduced branching followed by the death of the embryos between E14.5 and E17.5 (Kókai 2010).

3.1.6 Inhibiting c-Myc

Due to its central place in cancer (section 3.1.1), any drug that could target c-Myc transcription, the stability of the protein or its interaction with key protein partners is of extremely high therapeutic values. This is especially true since the over-expression of c-Myc is an aggravating factor and indicates a poorer prognosis in most cancers.

As such, the quest for c-Myc inhibitors has been extremely active (reviewed in Fletcher and Prochownik 2015 and Hui Chen, Hudan Liu, and Qing 2018). This quest has been made difficult by the fact that most of the c-Myc protein is unstructured. In this setting, the traditional computational biology and drug design approaches fall short because most of them rely on "molecular docking". In molecular docking, a virtual database of small chemical molecules is virtually "screened" against the structure of a known, structured protein. Good compounds usually dock in a "pocket" close to the active site of the protein, with the potential to disrupt the function of the protein. Since most of the sequence of the c-Myc protein has not been crystallized, this approach is not possible. Moreover, the bHLH domain of c-Myc that was obtained in (Nair and Stephen K. Burley 2003) revealed an extremely simple dimer, with a coil-coil that does not contain any targetable region.

To overcome this problem, several approaches were undertaken.

1. **Influencing the expression of the c-Myc gene.** Based on the existing knowledge on c-Myc promoters (reviewed for instance in Levens 2008), several approaches describe how to influence the transcription of c-Myc. (Kumar et al. 2008) identified an unstable structure, a G-quadruplex, in the P1 promoter, one of the main c-Myc promoters and stabilized ("trapped") it using synthetic modified oligonucleotides (Locked Nucleic Acids; LNA). Such a silencing approach was later generalized by the delivery of antisense LNA, a technique similar to siRNAs (Stein et al. 2010).

2. **Targeting the c-Myc protein or the c-Myc/MAX dimer** The most widely cited drugs that target c-Myc were identified in 2002-2003 and are named `10058-F4` and `10074-G5` (reviewed in Fletcher and Prochownik 2015). They were identified using a yeast-two-hybrid approach, in order to screen both for potent c-Myc inhibitors, but also for compounds that are cell-permeable and that specifically disrupt c-Myc/Max interaction. `10058-F4` has a 49 µM affinity whereas `10074-G5` had a 2.8 µM affinity.

   Compounds with higher affinity were later synthesized and screened. For instance, (Hart et al. 2014) designed a fluorescence polarization screen for the Myc/MAX interaction and identified `KJ-Pyr-9`, an inhibitor with a 6.5 nM affinity. In parallel, assays to screen compounds faster based on fluorescence were developed (Raffeiner et al. 2014).

   Another stream of research focused on the design of molecules bigger than the regular inhibitors, such as α-helix mimetics that interfere with the c-Myc/MAX binding interface (Jung et al. 2015).

   Finally, some authors tried to apply a more conventional screening approach by first using molecular dynamics to simulate the (unstructured) conformations of the c-Myc protein, and then try to find small-molecules that will stabilize one of these conformations, thus "trap" the protein in one conformation (C. Yu et al. 2016; Bayliss et al. 2017).

3. **Influencing c-Myc protein-protein interactions** Rather than directly targeting c-Myc or the obligatory c-Myc/MAX dimer, other researchers tried to specifically disrupt or stabilize given protein-protein interactions. The rationale behind these approaches is the folding-upon-binding mechanism: when an unstructured protein binds to one of its protein partner, it can undergo a folding-upon-binding transition to a well-folded, crystallizable protein complex. If one can derive a crystal structure of the protein complex, it is then possible to derive an inhibitor that will stabilize the protein complex, likely sequestering c-Myc in a potentially nonfunctional conformation.

   This approach was implemented by stabilizing the c-Myc/Aurora A complex using a drug (Gustafson et al. 2014; Richards et al. 2016), or by simply over-expressing HHex, a negative

regulator of c-Myc/Max (Marfil et al. 2015)

One more indirect study inhibited cIAP1, an E3 ubiquitin ligase that interact with a natural competitor of c-Myc with Max: MAD1. When MAD1 degradation is inhibited, it competes more with Max and Myc is ejected from its usual binding sites. This approach lead to the design of an efficient c-Myc indirect inhibitor (H. Li et al. 2018; Torchia and Ashwell 2018).

3.1.7 Biophysics of c-Myc in the nucleus

Several biophysical characteristics of c-Myc in mammalian cells have been determined, that help to understand how c-Myc could be involved in transcriptional regulation.

1. **A bulk vision.** First, despite having one degenerate binding site every 256 bp in average, not all the c-Myc motifs are equally occupied, and the main predictor of c-Myc binding is that the motif is located on the promoter of active genes (Nie et al. 2012; Lin et al. 2012). ChIP-seq indeed yields a few thousands binding sites (around 6000, depending on the analysis), which roughly corresponds to the number of active genes in mESC.

   Second, it is unlikely that all these binding sites are occupied 100% of the time in cells by c-Myc. Indeed the number of c-Myc protein in a single human (P493) cell has been estimated to be around 13000 (Lin et al. 2012) and one can determine (Izeddin, Récamier, et al. 2014) the bound fraction of c-Myc using live cell imaging, which was estimated to be around 10%, suggesting that even at the promoter of active genes, c-Myc could be very dynamic. Other estimates using quantitative Western Blots provide a similar order of magnitude, ranging from 6300 molecules per cell to 33000 in human lung fibroblasts (Rudolph, Adam, and Simm 1999). So far, mass-spectrometry-based techniques have not been investigated to quantify c-Myc (Vogel and Marcotte 2008; Zhou et al. 2013).

   Third, when c-Myc is progressively over-expressed during induction, the protein copy number can jump from 13000 to 360000 after 24h (Lin et al. 2012). In that case, the already bound sites show more intense peaks, and c-Myc starts to bind at so-called "lower affinity sites", such as enhancers.

   At the level of a whole organism, Fan-Minogue et al. 2010 used a split-luciferase system to monitor the interaction between c-Myc and the GSK3β kinase that phosphorylates c-Myc on Thr58.

2. **A single-cell picture of c-Myc.** At the single-cell level, few studies characterized the localization and dynamics of c-Myc. It has been long known that c-Myc is mostly a nuclear protein (with an estimated 4000 molecules in the cytoplasm and 29000 in the nucleus; Rudolph, Adam, and Simm 1999). At higher resolution, it seems that different phospho-isoforms of c-Myc tend to reside at different locations in the nucleus, with c-Myc-Ser62P being located at the periphery of the nucleus (Myant et al. 2015). This suggests an obligatory trajectory for c-Myc activation, in which every single protein will have to interact with the nuclear lamina in order to be activated. At super-resolution, we are not aware of any study looking at the nuclear organization of c-Myc.

   Finally, the question whether c-Myc might form highly heterogeneous clusters, or phase-separated droplets is still open, and *in vitro* experiments suggest that c-Myc can undergo phase separation under specific conditions (Boija et al. 2018).

3. **Dynamic picture of c-Myc.** Similar to many transcription factors, little is known about c-Myc dynamics. A first study published in 2009 (Sekiya et al. 2009) by the Zaret lab first suggested a link between the fast FRAP recovery of c-Myc and the fact that it does not act as a pioneer factor, which would explain its relatively low bound fraction.

   At the single-molecule level, c-Myc has been shown to exist in three sub-populations: an "immobile" fraction (9.5%), a slow fraction (diffusion coefficient D=0.5 µm²/s; 20.5%) and a fast fraction (13.5 µm²/s; 70%), and to explore the nuclear space in a relatively free, unconstrained manner, unlike P-TEFb (Izeddin, Récamier, et al. 2014).

3.2 Open questions in Myc biology

As of today, and despite decades of studies, a clear vision of how c-Myc regulates transcription is still missing, and many questions remain open. We propose an approach that could elucidate some of them in the next sections.

First, the determinants of the specificity of c-Myc for a given binding site are not known. Clearly, the presence of an E-box (`CACGTG` or a degenerate motif) seems to be necessary for c-Myc recruitment to a given locus, yet not sufficient (J. Guo et al. 2014). If chromatin accessibility is considered, then one can explain a little bit better the binding of c-Myc.

However, the preference of c-Myc for the promoters of active genes cannot be explained by sequence only, it is likely that protein-protein interactions are involved in the DNA specificity of c-Myc. This idea is supported by the fact that the deletion Myc-Box IV (MBIV) seems to impair DNA binding (Cowling et al. 2006). Furthermore, synthetic biology approaches have shown that the trans-activating domain of a TF can determine its sequence specificity (W. F. Lim et al. 2016). What are the relevant protein domains and/or protein-protein interactions that determine c-Myc DNA binding and dynamics?

Second of all, the kinetics of c-Myc binding on DNA are not known: how long does c-Myc stays on a promoter sequence? Does this time correlate with gene activation, as suggested by Clauß et al. 2017)?

Third, what are the post-translational modifications required for c-Myc activity? It seems that Thr58P and Ser62P are key for maintaining c-Myc stability (Sears 2004), but also SUMOylation (X.-X. Sun et al. 2018).

3.2.1 Approach

To understand the determinants of the dynamics of c-Myc, we decided to adopt the following approach, summarized in Figure 3.6:

- First, to build a cell line in which the c-Myc gene has been homozygously replaced by a tagged version amenable to single-particle tracking. To validate this cell line to show that it behaves indistinguishly from WT cells (section 3.3.1).
- Second, to build a series of constructs that could be transfected into the cells and that contain individual domain deletions of the main c-Myc domains (section 3.3.2)

As a readout, we initially decided to measure the following metrics:

- The diffusion coefficients of the various sub-populations of proteins using fast SPT
- The residence time of c-Myc on DNA using slow SPT

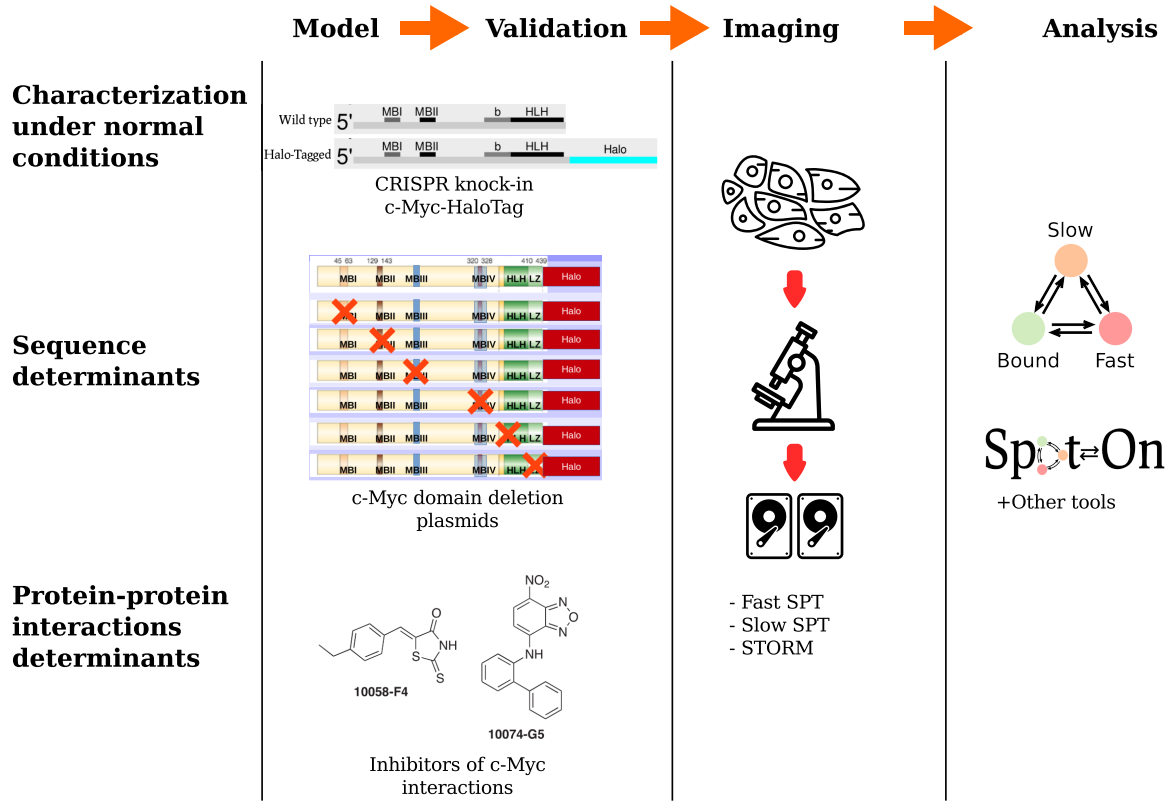


Figure 3.6: **Investigating the determinants of c-Myc dynamics**. Using three biological settings (*top row*: under "normal" conditions, in an endogenous context, *middle row*: when a protein domain of the c-Myc protein is deleted, in a transfection context, and *bottom row* when c-Myc interactome has been disrupted by inhibitors), our project aims at (*left column*) validating those models, then (*middle column*) to perform imaging (SPT, STORM) and (*right column*) to analyze the data using tools presented in previous sections.

Despite the fact that we only obtained preliminary results on this project, we made significant progress towards this goal, and in particular we validated most of the tools that could be used to answer these questions. We clarify some of the choices we made in section 3.2.2 and we discuss some of the difficulties we faced in the next sections.

3.2.2 Motivation

In this section we motivate a few of the technical choices made for this project.

1. **Choice of mouse embryonic stem cells (mESC).** First, we decided to conduct our experiments in mESC, a cell line in which c-Myc is expressed and essential. The choice of this cell line is motivated by the high number of genomic datasets published, the expertise of the Tjian-Darzacq lab in imaging and genome-editing this cell line. Furthermore, the choice of a non-cancer cell line was guided by the fact that mESC are likely to give more reproducible results. Finally, the role of c-Myc in the regulation of pluripotency justifies the use of mESC.

2. **Choice of the deletions.** Second, we decided to delete entire domains of the c-Myc protein. We decided to operate full-domain deletions as a first screen that could be refined later. This choice was motivated by the fact that many protein-protein interactions that have been mapped in c-Myc have been mapped within some of the Myc boxes (Figure 3.1 and 3.7). Our goal was to later refine the deletions, in order to more precisely target some specific protein-protein interactions.

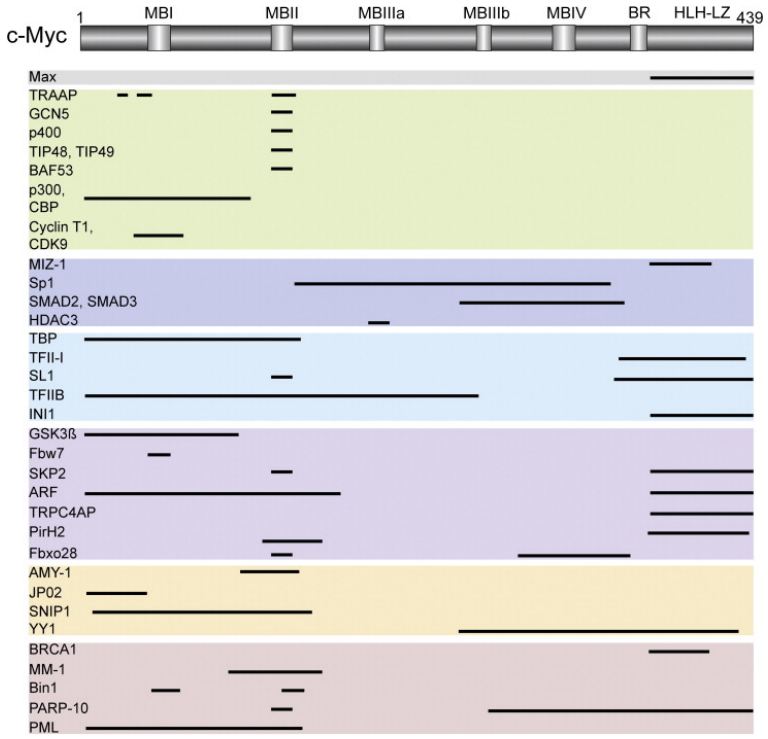


Figure 3.7: **Some of the mapped protein-protein interactions** Source: Tu et al. 2015

3. **Choice of the imaging technique** We decided to primarily use single-particle tracking (SPT) rather than other imaging or genomics technique. Our choice is motivated by the fact that SPT has been shown to be able to characterize subtle phenotypes in the dynamics of transcription factors. Such differences are expected to be difficult to highlight using traditional genomics such as ChIP-seq. Second, SPT provides a direct access to the dynamics of molecules, a key parameter to explain transcription regulation (section I.4).

3.3 Generation and validation of an endogenous knock-in cell line & plasmids

3.3.1 mESC cell line generation

1. **General approach.** In order to generate a model cell line amenable for imaging, and provided the dramatic effect of slight c-Myc over- or under-expression in cells, we opted to homozygously knock-in a tag (HaloTag Los et al. 2008; So, Yao, and J. Rao 2008, Dendra2 Gurskaya et al. 2006 or mEOS Paez-Segala et al. 2015) in mESC.

   This approach has several benefits: (1) the c-Myc-tagged protein will be under endogenous regulation, and should be expressed at endogenous levels. (2) Since all the c-Myc proteins are tagged, one can be reasonably confident that the tag does not alter the function of the protein, since no compensation will be possible with the wild type (WT) protein. (3) Genomics will be easier to perform than with regular transient transfections.

   CRISPR-Cas9 genome editing. We decided to use genome-editing techniques (Topaloglu 2005; Berdougo, Terret, and Jallepalli 2009; Urnov et al. 2010; Mata et al. 2012; Gaj, Gersbach, and Barbas 2013; Y. Kim et al. 2013; Mussolino et al. 2014), and more precisely CRISPR-Cas9 genome editing to replace the WT Myc gene by a tagged version. The CRISPR-Cas9 has become the tool of choice for targeted genome-editing, since a knock-in cell line can be derived in just a few weeks. The Cas9 protein targets preferentially all accessible regions (Horlbeck et al. 2016)

and is specifically targeted to a specific locus with a complementary RNA sequence, the sgRNA (Figure 3.8).

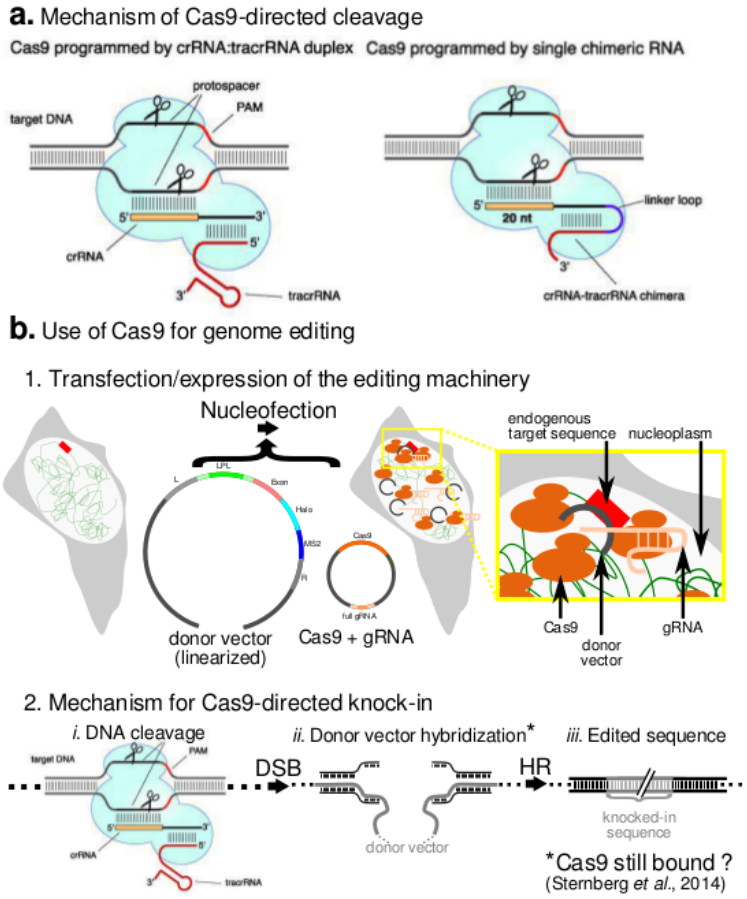


Figure 3.8: **Use of the CRISPR/Cas9 for genome editing**. Parts of the figure are derived from (31), figure 5. Source: Woringer 2014

Design of the vectors. We first used online tools to design the sgRNA sequence (`http://crispr.mit.edu`, not online anymore, but other tools exist, Spahn et al. 2017). We designed two sgRNA sequences and cloned them into two separate vectors that also contained the Cas9 sequence. Second, we used Gibson cloning (Gibson et al. 2009) to design a repair vector that contained $\sim$ 500 bp homology for the 3'-end of the MYC gene and the sequence of the tag to insert. These vectors were then transfected into mESC and we proceeded to FACS sorting and clonal selection. The editing process is pictured in Figure 3.9).

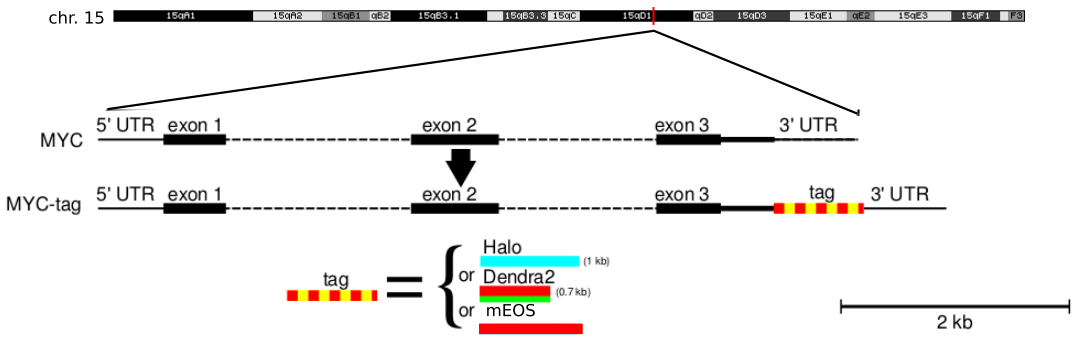


Figure 3.9: **Targeting the MYC gene**. (top). location of the MYC gene on chromosome 15 and detail of the three exons, four introns MYC gene. (bottom). MYC structure after editing.

2. **Choice of sequences.** We decided to clone the tags in C-ter of c-Myc to avoid disruption of any known regulatory function of c-Myc (as reviewed in section 3.1), including potential interaction with the various promoters and alternative transcription start sites present in the locus. Furthermore, a previously described C-ter fusion has been published (c-Myc-GFP; Nie et al. 2012). Note, however, that tagging in the N-ter of the c-Myc protein seems also viable, and a GFP-Myc mice has also been published (C.-Y. Huang et al. 2008).

   Furthermore, a tag in N-ter is hard to characterize, because if it disrupts protein-protein interactions, it is likely to disrupt poorly characterized interactions between MBI and unknown proteins. Conversely, the tagging strategy we chose places the tag close to the DNA-binding domain, bHLH of c-Myc. It is easy to check with high sensitivity by ChIP-seq to what extent the binding profile of c-Myc is affected by the tag: it is easier to check for DNA binding defects than for (uncharacterized) protein-protein interaction defects.

   In terms of tags, we initially opted for a HaloTag (Los et al. 2008; So, Yao, and J. Rao 2008), since it can be conjugated to a wide variety of photostable fluorophores of different colors and properties. After a first unsuccessful genome-editing attempt, we also generated constructs and cell lines with a different linker, and with two different tags: Dendra2 (Gurskaya et al. 2006) and mEOS3.2 (Paez-Segala et al. 2015). Both mEOS and Dendra2 are photo-convertible proteins. They are less bright and less photostable than synthetic dyes conjugated to HaloTags, but should guarantee that the protein fusion is not fluorescent anymore when c-Myc gets degraded, which is not the case with organic dyes that might remain in the cell long after the degradation of the fusion protein. A summary of the genotypes derived is presented in Figure 3.10.

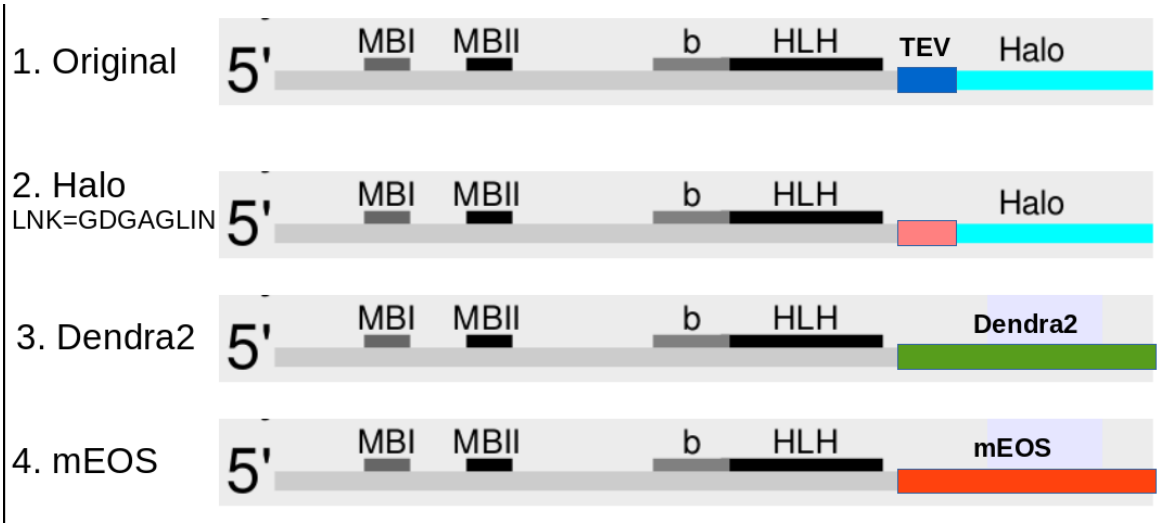


Figure 3.10: **Summary of the genotypes of the cell lines generated**

3. **Cell line selection.** We first transfected the c-Myc-TEV-HaloTag construct (Figure 3.10), together with one (sgRNA, Cas9-Venus) construct. After one day, the Venus-positive cells (approx. 10%) were FACS sorted in order to enrich for transfected cells. 48 h later, the cells were labeled with the TMR-Halo dye and FACS-sorted, in order to keep only the HaloTag-positive cells. These cells were plated at low density and colonies were left to grow for 7-15 days. When colonies were big enough to be visible, we picked $\sim$ 96 of them and genotyped them using direct PCR.

   To our surprise, no cell was above the FACS threshold, suggesting either an issue in the genome-editing, or that c-Myc levels were too low to provide a good contrast. We thus kept the 0.1% of the cells that had the highest fluorescence (approx 5000 cells) and plated them as described above. Most of the genotyped clones were heterozygous, and the homozygous clones displayed a slower growth rate and abnormal morphology.

   We thus generated three additional transfections, with the c-Myc-GDGAGLIN-Halo construct (c-Myc-HaloTag with a different linker), with c-Myc-Dendra2 and c-Myc-mEOS. In all cases, and as with the previous experiment, we obtained very few cells above the FACS sorting threshold (no cell above the FACS threshold for the c-Myc-Halo line, $\sim$ 600 cells for Dendra2/mEOS). Of these cells, we derived very few colonies (12 colonies for Dendra2, 7 colonies for mEOS). Nonetheless, most of them were properly edited (Figure 3.11), and we obtained homozygous and heterozygous clones for all the constructs.

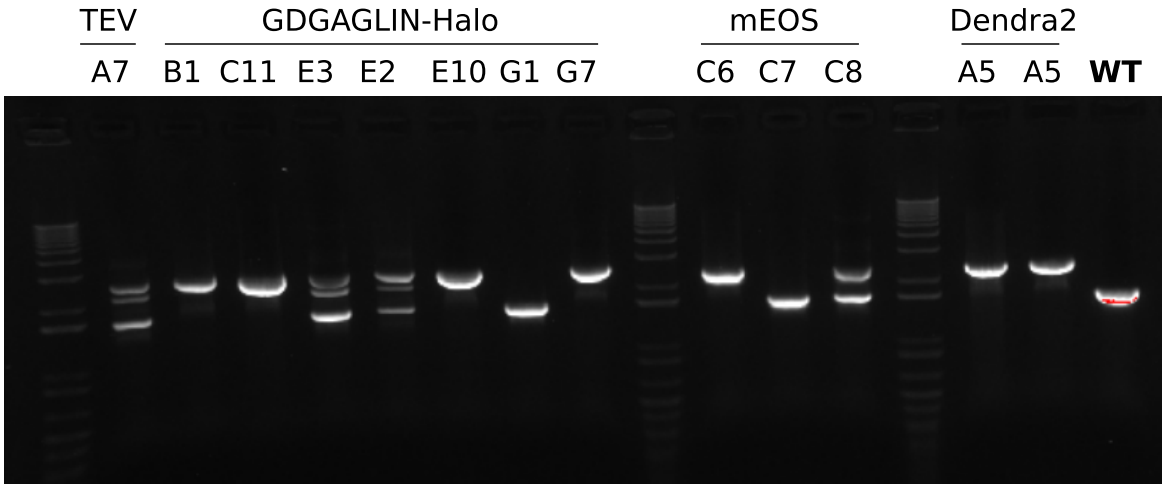


Figure 3.11: **Genotyping of several clones using a genomic PCR.** Four constructs were transfected: cMyc-TEV-Halo (TEV), cMyc-GDGAGLIN-Halo (GDGAGLIN-Halo), cMyc-mEOS (mEOS) and cMyc-Dendra2 (Dendra2). They are compared to the wild type cell line (right lane; WT). Non-edited cell lines appear as WT, heterozygous lines appear with two (or more) bands, homozygous with a single band running slower than the WT. The MYC locus of the cell lines was further sequenced.

The MYC locus of the homozygous clones was sequenced, and the insertion was identical to the repair vector. The cells were checked for mycoplasma contamination (Figure 3.12). At this step, we had derived homozygous clones with c-Myc tagged with a HaloTag (with two different linkers), with Dendra2 and with mEOS.

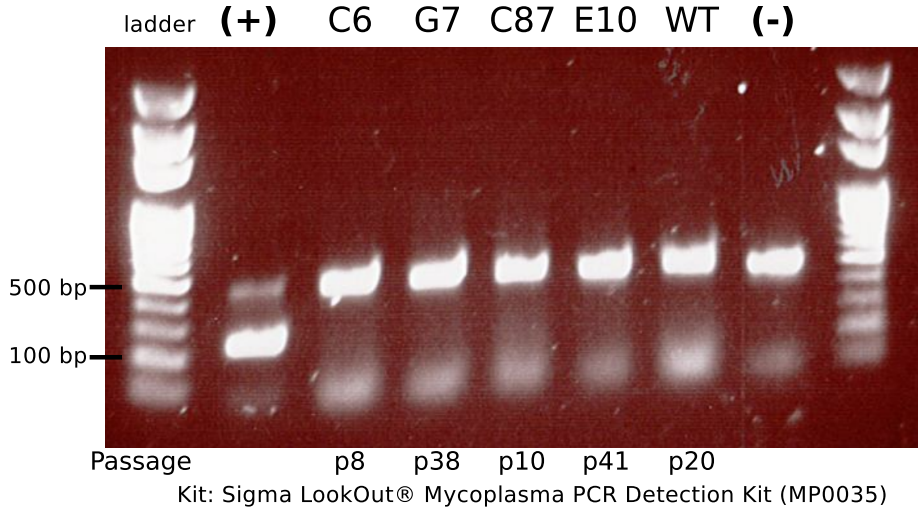


Figure 3.12: The selected clones are negative for Mycoplasma contamination.

We then observed the cell for morphology and in fluorescence. The c-Myc-Halo clone displayed a normal morphology (Figure 3.13-a-center). On the other hand, the only homozygous c-Myc-Dendra2 clone displayed an abnormal phenotype, resembling differentiated or neuron-like cells (Figure 3.13-a-left). This clone was not used for subsequent analysis. Finally, the mEOS cells also displayed a normal morphology in transmission light (Figure 3.13-a-right). However, when observed in fluorescence, the c-Myc protein displayed a totally abnormal localization: the c-Myc protein is supposed to be relatively uniform in the nucleus, and not focused in puncta, as displayed in Figure 3.13b. We then decided to only use the c-Myc-Halo cell line, since we derived independent clones for this construct.

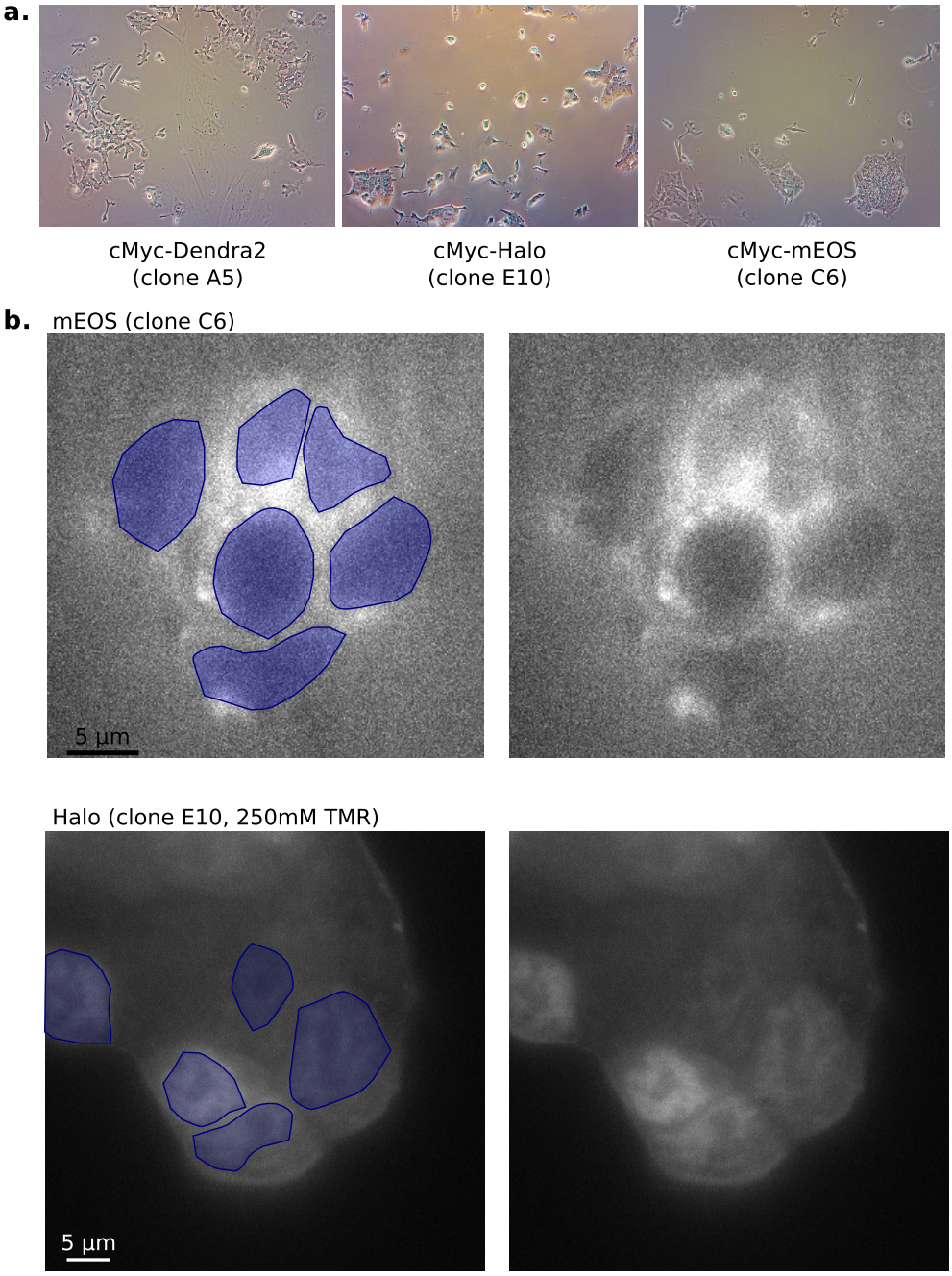


Figure 3.13: **Morphology of three clones of edited cells. (a)** As observed in transmission light microscopy. **(b)** c-Myc mEOS and c-Myc-GDGAGLIN-Halo (E10) protein localization. c-Myc-Halo was labeled with 250 nM TMR. The nuclei were hand-drawn for readability. (In this plot careful the scale bars differ and also the exposure times).

3.3.2 Transient transfection approach

1. **General approach.** In order to build constructs that could be use to assess the effect of domain deletions in the c-Myc protein, we generated a series of constructs in which we individually deleted the main domains of c-Myc: MBI, MBII, MBIII, MBIV, bHLH and LZ. For all these constructs, we initially generated a HaloTag, a mEOS and a Dendra2 version.

   We chose to place the c-Myc-tag sequence under a ribosomal protein L30 promoter (L30 promoter), in order to ensure mild expression of the vector, and also because the traditional viral, CMV promoter is methylated and silenced in mESC (Herbst, Ball, et al. 2012).

In practice, we used a combination of traditional cloning and Gibson assemblies to generate the vectors. The vectors also included a second ORF, EGFP-nuc. This second construct will be used to quickly screen for transfected cells in the case where photo-activatable dyes, that reside in the dark state without photo-activation, are used in combination with HaloTagged proteins. An example vector is presented in Figure 3.14a.

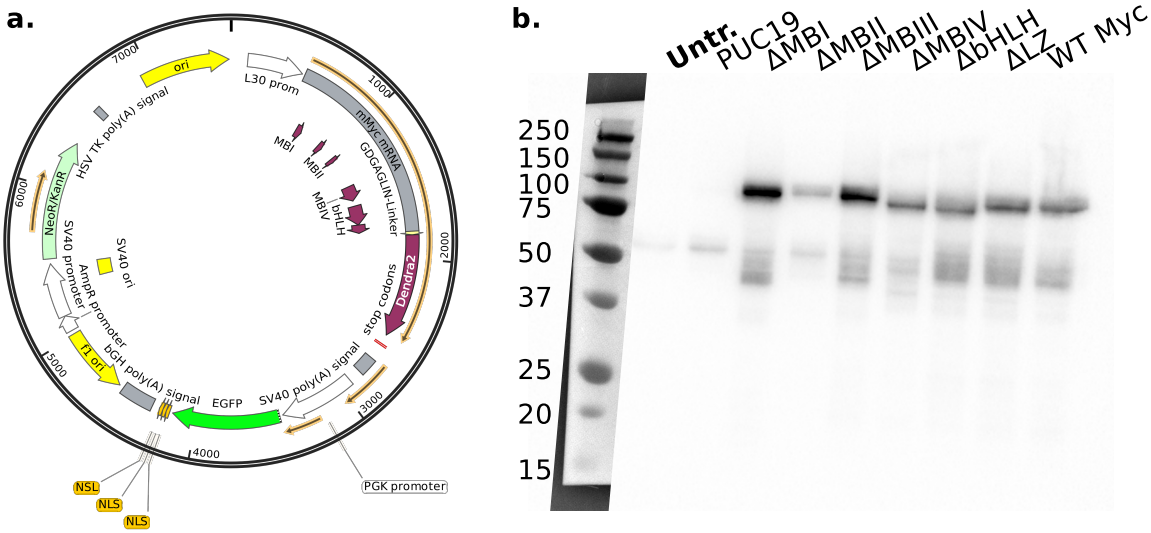


Figure 3.14: **(a)** Sample plasmid map of a c-Myc-Dendra2 construct. **(b)** Western Blot of transfected constructs in 293T cells using an anti-c-Myc antibody.

2. **Plasmids generated.** All in all, we generated the 15 constructs, summarized in Table 3.1.

Table 3.1: The numbers correspond to the identifier in the plasmid library. Due to the localization issue of the mEOS cell line, we prematurely interrupted the construction of the mEOS library.

Domain deletion v	Tag > Halo	mEOS	Dendra2
none (WT)	#76		#91
MBI	#77		#92
MBII	#78		#93
MBIII	#79		#88
MBIV	#80		#90
bHLH	#81		#94
LZ	#82	#100	#89

3. **Plasmid functional validation.** As a first validation, we verified the expression of the constructs by transfecting them into the easily transfectable 293T cells and performed a Western Blot against c-Myc (Figure 3.14b). All the constructs expressed reasonably well, which justified their use in further experiments.

3.3.3 Cell line controls

1. **General approach.** Once the cell lines and constructs were built, we aimed at validating them. More specifically, we wanted to make sure that the cell line we designed was only minimally affected by the addition of the tag on the c-Myc protein, that the protein expressed properly and that the cells were amenable to live cell imaging. In this section and the following, we compared the characteristics of the non-edited mESCs with those of the c-Myc-tag cell lines. Our goal was to prove that the cell line we generated is amenable to single-cell molecule imaging.

We assessed that the cells still displayed regular mESC markers (section 2), and we tried to assess that the expression of c-Myc target genes was not affected by the addition of the tag (section 3).

2. **Pluripotency genes expression.** As a second control (after morphology assessment, Figure 3.13), we verified that the cells still expressed key pluripotency markers: Oct4, Nanog, Sox2. Non genome-edited cells and genome-edited cells were grown on coverslips and we performed immunofluorescence (IF) on the samples with specific antibodies and DAPI staining. The cells were then imaged under similar conditions and the average nuclear fluorescence level was quantified using Fiji (Schindelin et al. 2012).

   We decided to opt for IF characterization rather than a more conventional qPCR because IF not only gives access to the expression level of the pluripotency genes, but also to more subtle phenotypes (change in localization and cell-to-cell variability, documented in Clavería et al. 2013; Díaz-Díaz et al. 2017) that are not usually assessed using qPCR. However, it should be noted that IF-based imaging approaches usually suffer limitations when one tries to characterize subnuclear patterns. These artifacts are most of the time due to the fixation protocol (Huebinger et al. 2018; Bedino 2003).

   Overall, the cells expressed properly the pluripotency markers, with an expression pattern that matched the WT control (Figure 3.15, left column). However, compared to the WT reference, the c-Myc-Halo G7 clone had a reduced expression of the (Oct4, Sox2, Nanog, c-Myc) proteins. Even though a precise quantification is difficult when using IF, it seems that the edited clone has a 30-50% lower expression of the pluripotency markers compared to WT. A precise quantification of the fluorescence intensity levels would be interesting, and automated segmentation tools based on the DAPI signal could greatly help in this direction.

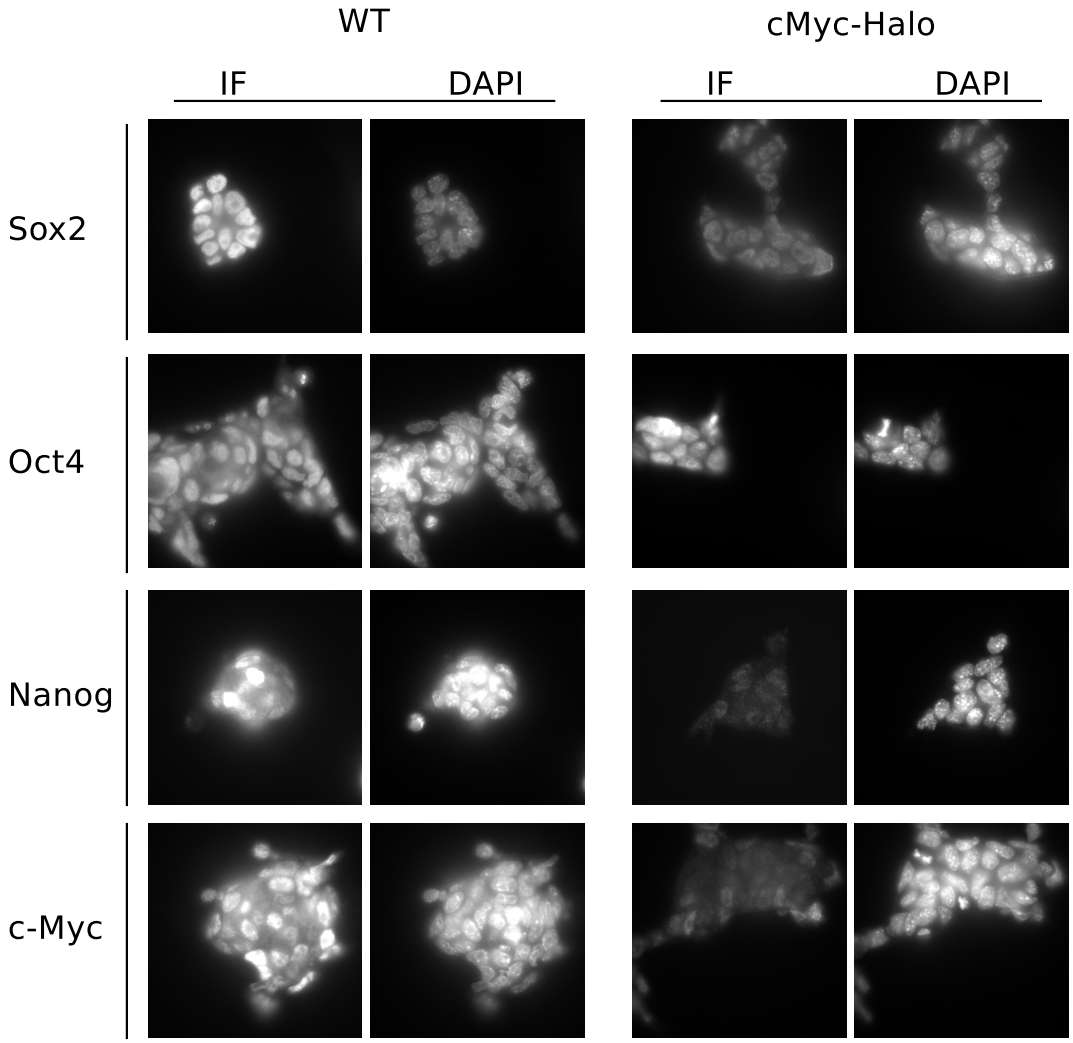


Figure 3.15: Immunofluorescence of the Sox2, Oct4, Nanog and Myc pluripotency factors in wild type cells (left) and a c-Myc-Halo clone (right).

3. **RNA-seq.** Having demonstrated that the genome-edited still expressed common pluripotency markers, although at reduced levels, we set out to determine whether the transcriptional program differs between the WT and the genome-edited cells by RNA-seq of the messenger RNAs.

   We prepared triplicate samples of the WT and the genome-edited cells and harvested them on three different days using the triZOL reagent. Total RNAs were then purified, quality-controlled using the BioAnalyzer and sent to the sequencing facility, who prepared the libraries and performed NGS. As of today, the sample have not come back from the sequencing facility.

3.3.4 Protein expression

1. **General approach.** After verifying that the cell line displayed the expected embryonic stem cell phenotypes by morphology inspection, immunofluorescence (section 2) and possibly RNA-seq (section 3), we focused more on the characteristics of c-Myc expression.

   Indeed, in order to derive reliable measurements on c-Myc dynamics, one needs first to check that the background is not modified, and then that the properties of the protein are not measurably changed. In order to perform single-molecule imaging, focused on measuring the following parameters:

   - The expression level of c-Myc (section 2)

- The interaction with c-Myc's main interacting partner, MAX (section 3)
- The lifetime of the protein (section 4)
- Protein-DNA interactions (section 5)

2. **c-Myc expression level.** First, we verified that the tagged version of c-Myc expressed properly. We performed a Western Blot (WB) using an anti-c-Myc antibody, and paid attention to existing degradation products that might indicate the presence of free HaloTag proteins (unfused to the c-Myc protein), Figure 3.16. This blot was realized with the help of Claire Darzacq.

   Despite the uneven loading of the gel, one can make a few remarks. (1) Compared to the actin control, the c-Myc protein seems to express at comparable levels in WT and cMyc-Halo cells, although a quantitative WB is needed to provide definitive conclusions. (2) The c-Myc antibody stains a single (doublet) band, suggesting that no major degradation occurs in the HaloTag. (3) Each band appears as a doublet, that is thought to represent the two isoforms of c-Myc (Myc1 and Myc2, section 2). These two bands are present in the c-Myc-Halo cell line, suggesting that the translational regulation of c-Myc has not been affected.

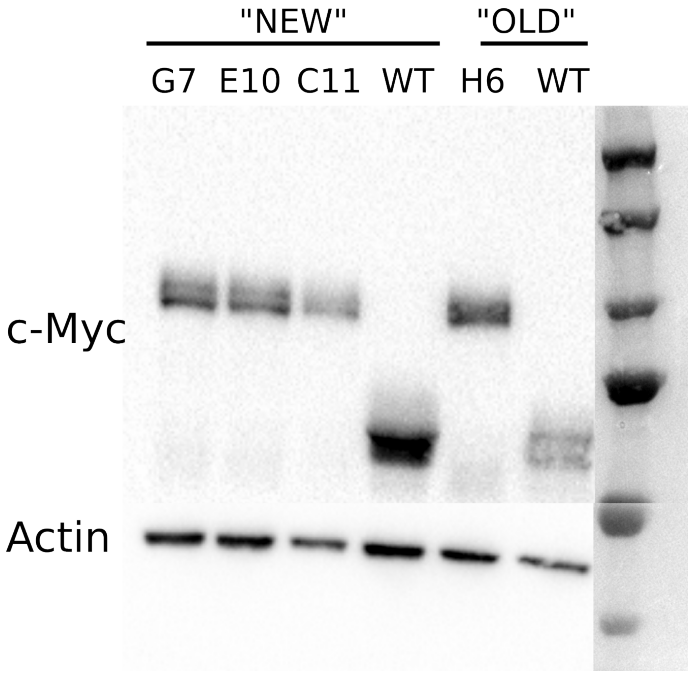


Figure 3.16: Western Blot of several genome-edited c-Myc lines. WT represent the Wild type lanes. The WT c-Myc runs around 60 kD, the c-Myc-Halo fusion around 150 kD.

A second, complementary question remains, whether there exists in cells a fraction of partially degraded c-Myc proteins that will not be detected by the c-Myc antibody we used (Abcam's ab32072, that recognizes c-Myc N-ter region, on amino-acid 69), but will still carry a partially degraded or full-length HaloTag. Such products are unwanted because they correspond to degraded c-Myc proteins but will contribute some fluorescence signal until the Halo-ligand fluorophore is degraded.

We tried to address this question by performing a fluorescence-based Western Blot. The cells were labeled using TMR-Halo and incubated for zero or one hour. We then tried to reveal the blot using a fluorescence scanner. The bands should have provided us with an idea of the fractions of fluorescent peptides present in a cell. Unfortunately, this experiment did not yield conclusive results. We thought about relying on the characterization of full-length HaloTag, and to perform a WB against the HaloTag, using an antibody (Promega's G928A) targeting the HaloTag peptide, but did not have time to perform the experiment.

3. **Interaction with MAX.** Second, we wanted to check whether the interaction between c-Myc and MAX was preserved, despite the tag being close to the c-Myc/MAX interface (the bHLH region). Even though it appears unlikely that the cells will retain an embryonic stem cell phenotype without a properly functioning c-Myc/MAX interaction, we decided to check whether c-Myc co-immunoprecipitates (Co-IPs) with MAX.

   This experiment was performed in 2014 with the help of Claudia. c-Myc fusion proteins were transfected into N-Tera2 cells (Haile et al. 2014), and the pull-down was performed using an anti-c-Myc antibody. The pull-down was then blotted against c-Myc and MAX (Figure 3.17). Because these experiments were performed very early compared to the genome-editing process, they were performed using a different cell system (NT2 cells), and different tags (YFP, GFP and Dendra2). These results indicate that the addition of several tags does not disrupt the interaction with Max.

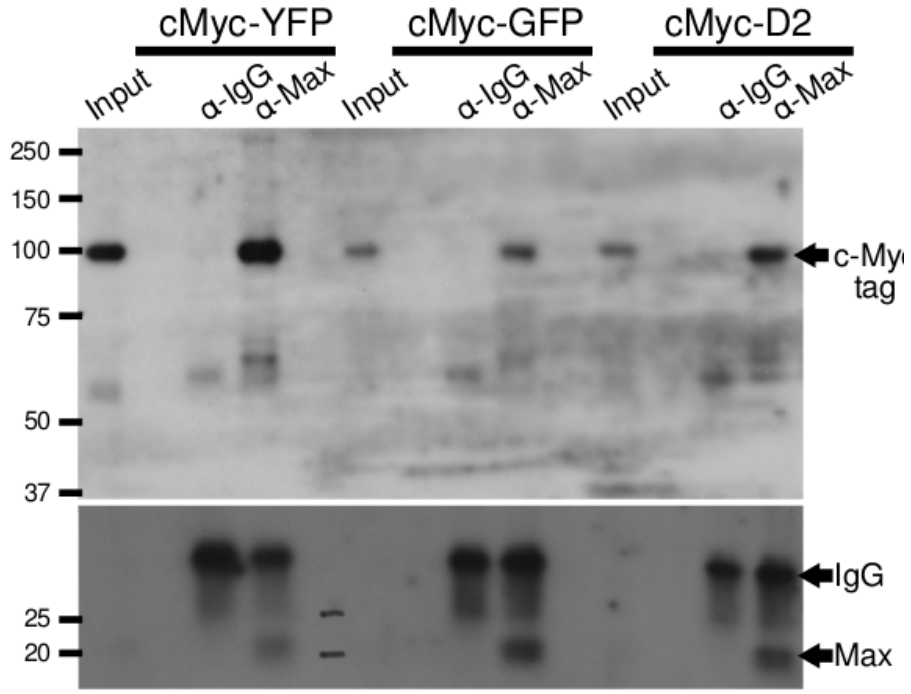


Figure 3.17: c-Myc-YFP, c-Myc-GFP and c-Myc-Dendra2 all co-immunoprecipitates with MAX.

4. **c-Myc protein lifetime.**

   (a) *Lifetime changes with respect to WT.* Third, we assessed whether the lifetime of the c-Myc protein was changed by the addition of the tag. To do so, we performed a standard protein lifetime assay using a cycloheximide block (Kao et al. 2015). Cycloheximide is a fast inhibitor of protein synthesis. When the cells are treated with cycloheximide, no new protein synthesis occurs, and one can follow the degradation rate of a protein using a time course of Western Blots.

   We performed a cycloheximide block using previously published concentrations for mESC (Tichy et al. 2012) and collected samples of the WT cells and the genome-edited cells at various time points. We assessed the lifetime of the c-Myc-HaloTag and the c-Myc-mEOS cell lines over a 6-hours time course. The results are given in Figure 3.18. No clear difference emerged in the lifetime of the tagged proteins versus WT.

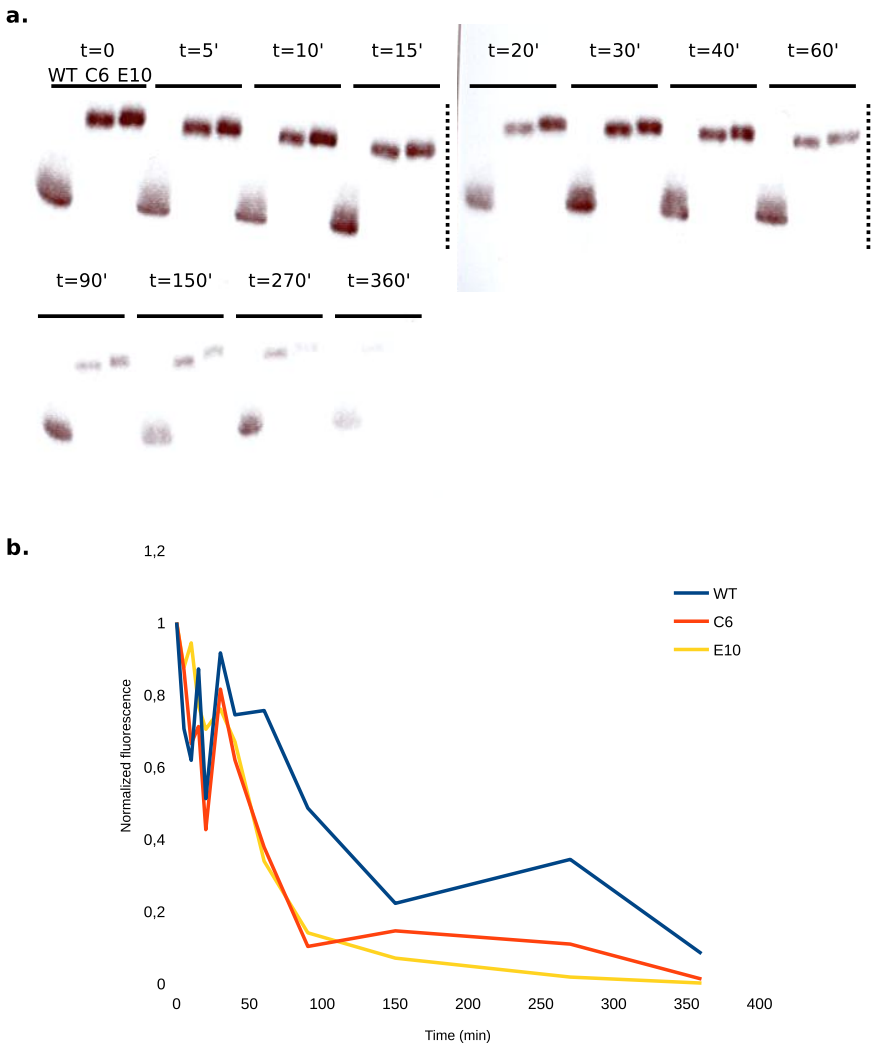


Figure 3.18: **Estimate of c-Myc lifetime using a cycloheximide block. (a)** Raw time course **(b)** Quantification of the gel.

(b) *Half-life of the c-Myc protein.* We then compared the half-life of the proteins, based on the quantification of the WB (Figure 3.18). The WT c-Myc appeared to have a half-life around 90 minutes (higher than what has been previously proposed, where c-Myc is known to have a 20-30 minute half-life), whereas both tagged cells exhibit a $\sim$ 45' half-life. Nonetheless, these results were not replicated so far and await confirmation.

All-in-all, these results also suggest that experiments using a Halo ligand should be conducted within 30-40' after labeling, in order to avoid that a too high fraction of the dye ends up in degraded forms of of the protein.

To complement this WB-based approach, we performed time-lapse, live-cell imaging (Figure 3.19. We observed that $\sim$ 20 minutes after the beginning of the imaging, the dye (and presumably the c-Myc protein) started to accumulate in perinuclear bodies that could be lysosomes or degradation bodies. This observation further reinforced the need to proceed with imaging relatively quickly after labeling.

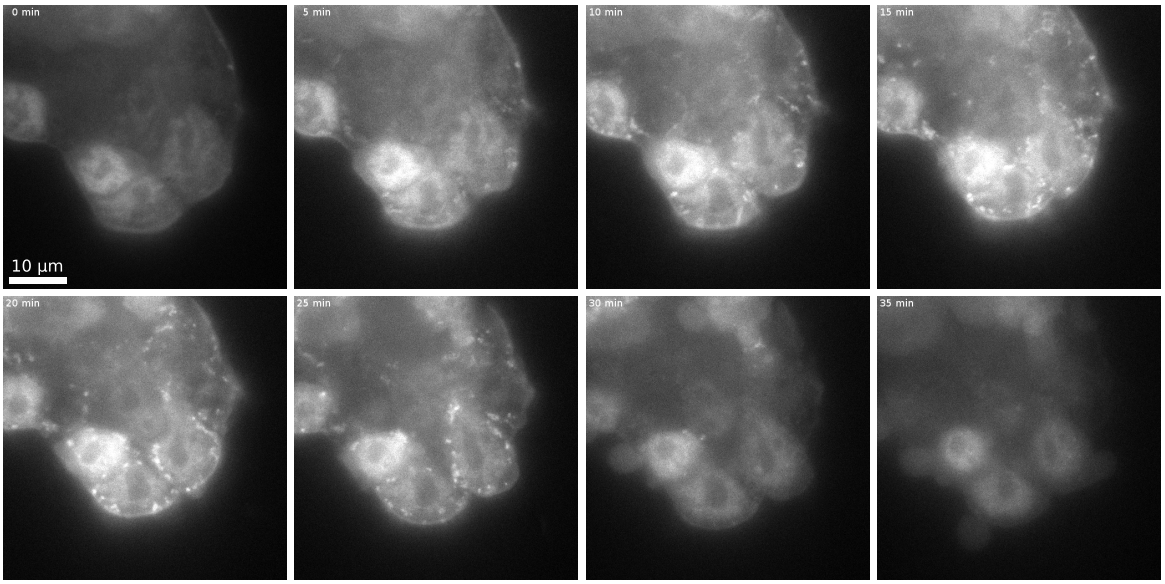


Figure 3.19: Degradation of the c-Myc protein over time, observed by labeling the c-Myc-Halo cell line E10 with 250 nM TMR. Cells were exposed for 150 ms every minute. Cells die after 35-40 minutes. One screenshot every 5 minutes

5. **Protein-DNA interactions.** Then, we aimed at determining whether the tag perturbed c-Myc binding pattern to DNA by ChIP-seq. The goal was to determine whether the binding pattern of c-Myc might be differentially affected depending on the locus, reflecting for instance the impairment of the interaction with one recruitment-related partner. This experiment is particularly important because the tag is located close to the bHLH DNA-binding domain.

   ChIP-seq protocol. We used a previously validated c-Myc antibody to perform ChIP-seq (Chronis et al. 2017), that was also validated using qPCR by Inma Gonzalez (Navarro lab) and gave a "not very strong signal, but still ok". For each cell line tested, we performed three biological replicates. For each replicate, we grew a few million cells and followed standard ChIP-seq protocols (Testa et al. 2005; Carey, Peterson, and Smale 2009; Ma and W. H. Wong 2011) to obtain purified chromatin samples. We decided not to perform a spike-in control, and thus only compare relative binding at one locus (K. Chen et al. 2016).

   qPCR design. We designed qPCR primers to validate the efficiency of the ChIP. We focused on two well-characterized c-Myc target genes: ODC1 (ornithine decarboxylase) and NPM1 (Bello-Fernandez, Packham, and Cleveland 1993; Zeller, XiaoDong Zhao, et al. 2006 and section 3.1.4. The primers were designed based on previously published ChIP-seq (Chronis et al. 2017), cross-referenced with existing qPCR publications. For each of these genes, a primer was designed on the corresponding ChIP-seq peak, and another one in a control region (Figure 3.20).

   Due to the imaging issues presented in section 3.4, the collected samples have not been sent for sequencing yet.

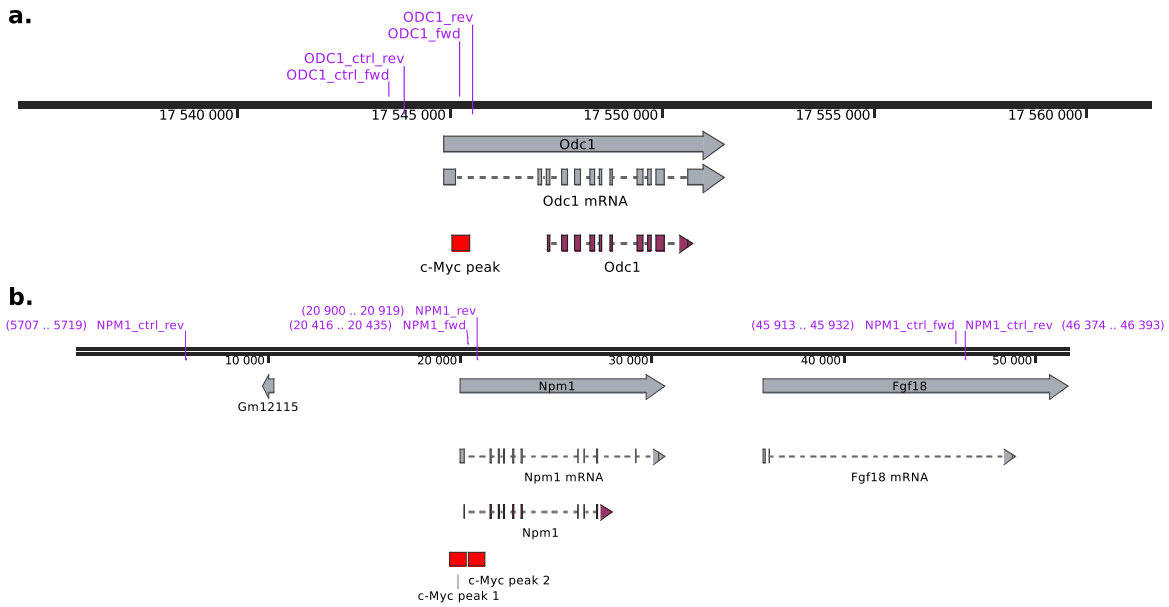


Figure 3.20: Location of the qPCR primers to validate the antibody and the ChIP. The location of the primers is highlighted in purple. The location of the c-Myc binding peaks is presented with red boxes.

3.3.5 Protein number characterization

The quantification of the absolute number of proteins per cell is a key biophysical parameter, but it is usually complicated to derive. For instance, a popular approach involves the purification of the protein of interest followed by an absolute *in vitro* titration, followed by a quantitative Western Blot. Another approach, used in the quantification of the Bicoid gradient in *Drosophila*, relies on the purification of fluorescence protein, and to incubate the cells in a medium that contains a given concentration of the fluorescent protein (Morrison et al. 2012).

These two approaches rely on complex protein purification techniques. Recently, a FACS-based technique that uses a calibrated cell standard was proposed to quantify the expression level of fluorescently tagged, endogenous proteins (Cattoglio et al. 2018). In this approach, one compares the expression level of a tagged, knocked-in proteins such as our c-Myc-Halo cell line with a reference cell line with the same tag (such as CTCF-Halo or Rad51-Halo), in which the absolute protein number has been determined using a biochemical method. Once grown in the same conditions, the cells are labeled similarly and analytical FACS is performed for the two cell lines. Then, the ratio of the mean fluorescence gives the ratio of mean protein numbers between the two cell lines. Since the number of proteins per cell is known in the reference cell line, one can retrieve the protein copy number in the sample cell line (Figure 3.21. The authors provided us with the reference cell lines.

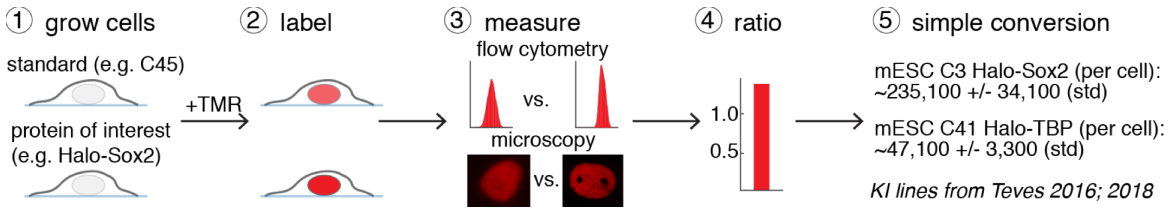


Figure 3.21: (1) Grow cells expressing the Halo-tagged protein of interest together with one of the standards described here (e.g. C45; Figure 1B). (2) After labeling with a fluorophore (e.g. TMR or a JF-dye), the relative fluorescence intensity can be measured using either flow cytometry or microscopy (4) and thus the absolute abundance calculation (5).(source: Cattoglio et al. 2018)

We followed the procedure proposed by Cattoglio and colleagues and quantified the absolute protein

number in two independent c-Myc-Halo clones and we find that there are approximately 61000-78000 c-Myc molecules per cell ($\sim$ 25 minutes after washing the medium), Figure 3.22. However, we only performed the FACS quantification once, and thus do not have reliable error estimates on this number.

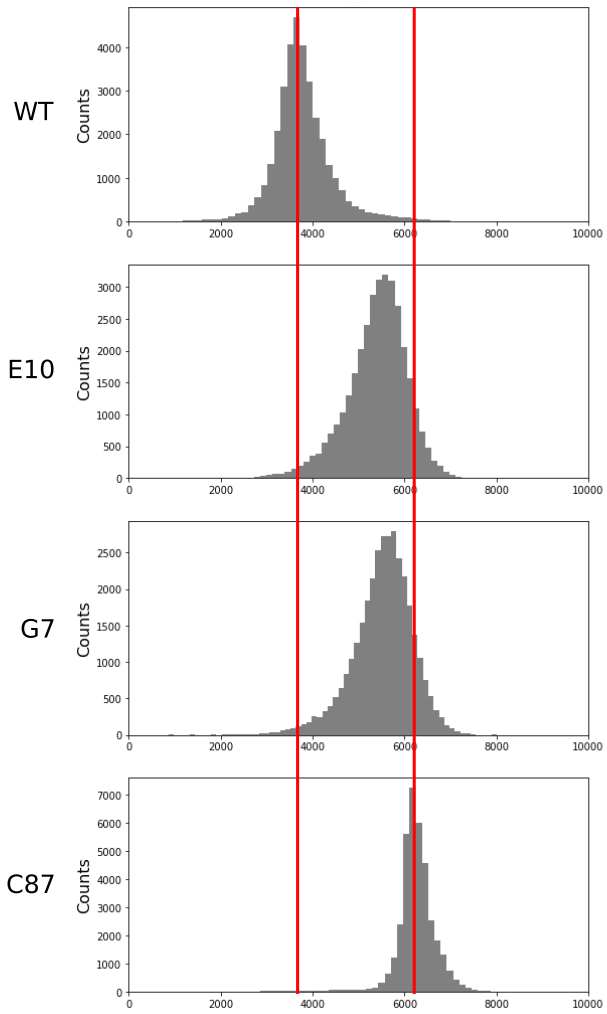


Figure 3.22: Distribution of the fluorescence intensity of TMR labeled cell lines, assessed by analytical FACS. The cell lines observed are mESC, either WT (top row), a c-Myc-Halo knock-in (two middle rows, clones E10 and G7) or a Halo-CTCF cell line (bottom row).

3.3.6 Inhibitors

We screened a part of the literature and the ChEMBL database for c-Myc-interacting compound, yielding $\sim$ 60 compounds. The results were then manually selected based on the quality of the inhibitor and other characteristics. We then focused on the commercially available compounds, and ordered some of them.

We didn't have the time to validate the c-Myc inhibitors that we initially planned to use for the imaging experiments.

3.4 Preliminary SPT experiments

3.4.1 Reanalysis of previously published data

In order to validate our own SPT data, Ignacio Izeddin kindly provided us with c-Myc-Dendra2 SPT data acquired in U2OS cells at fast frame (10 ms/frame). The authors had derived a kinetic modeling approach very similar to ours (Izeddin, Récamier, et al. 2014). Using Spot-On, we compared the jump length distribution with fitted theoretical model and obtained a reasonable fit (Figure 3.23. Table 3.2 compares the fitted values in the original publication and in our reanalysis. We consider that these results are globally concordant, and validate that our analysis pipeline will produce results that can be compared with existing literature.

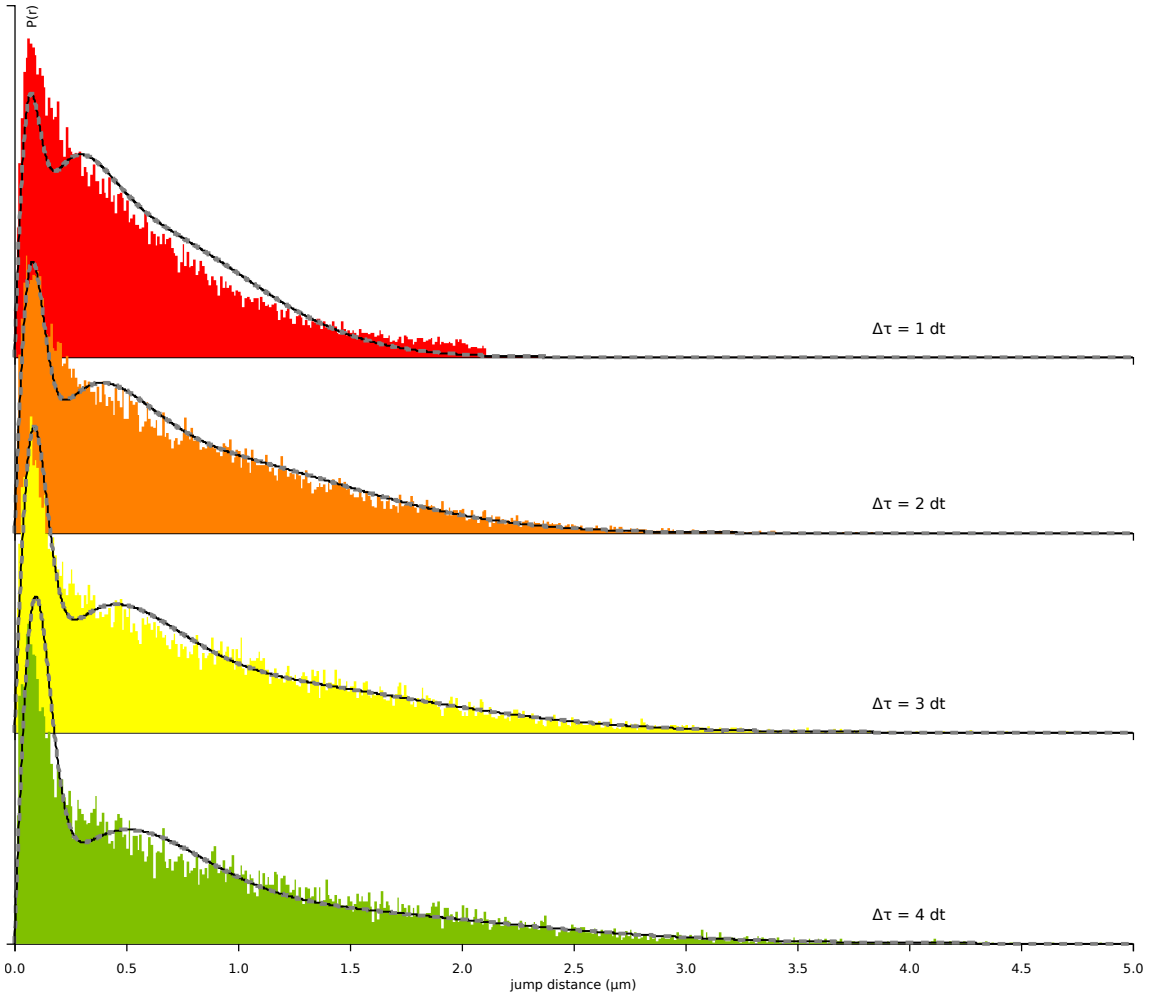


Figure 3.23: Jump length distribution and corresponding model fit of the data published in Izeddin, Récamier, et al. 2014.

Table 3.2: Analysis (left) and reanalysis using Spot-On (right) of the dataset published in Izeddin, Récamier, et al. 2014.

Parameter	Izeddin et al., 2014	Spot-On reanalysis
D_{bound} (F_{bound})	<0.1 μm²/s (10%)	< 0.1 μm²/s (6%)
D_{slow} (F_{slow})	0.5 μm²/s (20%)	2.7 μm²/s (20%)
D_{free} (F_{free})	13.5 μm²/s (70%)	17.9 μm²/s (74%)

3.4.2 Preliminary imaging in UC Berkeley

Once we obtained and reanalyzed existing c-Myc SPT data, we decided to perform preliminary SPT imaging. The cells were labeled using PA-JF_{549} and imaged at high frame rate (6.5 ms/frame) using stroboscopic illumination (2 ms). The recorded trajectories were then analyzed using Spot-On (Figure 3.24).

Our preliminary results are summarized in Table 3.3. They are surprisingly close to our reanalysis of the data by (Izeddin, Récamier, et al. 2014), suggesting that our model behaves similarly as the one developed previously, despite the change in cell line (U2OS vs. mESC), labeling strategy (transfected c-Myc-Dendra2 vs. homozygous knock-in c-Myc-Halo) and acquisition setting (different microscopes, exposure times and illumination schemes).

Given these encouraging results, we decided to import in Paris the SPT imaging approach, and use this preliminary data as a reference point to optimize the setup in Paris.

Table 3.3: Parameters extracted from the preliminary SPT experiment presented in Figure 3.24.

Sub-population	D ($\mu m^2/s$)	Fraction
Bound	<0.1	4%
Slow	2.9	31%
Fast	12	65%

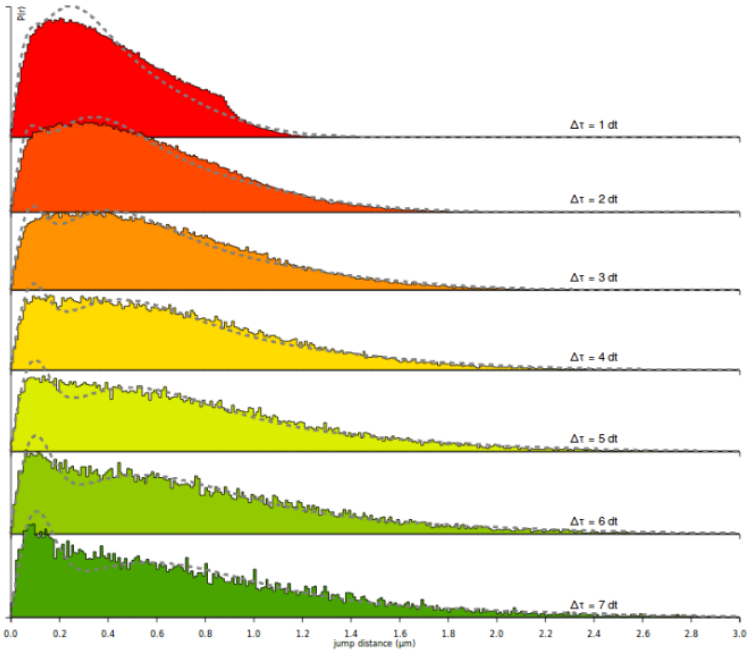


Figure 3.24: Spot-On analysis of SPT of c-Myc performed at 6.5 ms/frame. The extracted coefficients are presented in Table 3.3.

3.4.3 Preliminary imaging in Institut Pasteur

1. **The microscope.** Despite the strong experience of the Zimmer lab in single-molecule imaging, to move from PALM/STORM imaging to SPT requires a few adaptations to the microscope, that we conducted jointly with Mickaël Lelek.

   First of all, we installed a 37°C+CO_2 incubator stage on the microscope and calibrated it, in order to keep the cells alive for an extended period of time.

Second, after preliminary experiments, we decided to recollimate the laser beam in order to illuminate a $\sim$ 15x15 µm² region rather than a 50x50 µm² region, in order to multiply the incident power per area by a factor of $\sim$ 10. This operation was performed by Mickaël and necessitated to create an alternative light path, so that the microscope could be used in both "large illumination" and "narrow illumination" modes.

Third, we rewired the AOTF system so that stroboscopic illumination (J. Elf, G.-W. Li, and X. S. Xie 2007) was synchronized to the camera, allowing to illuminate the sample with 405 nm laser during the transfer ("dark time") of the camera, and to pulse the excitation laser during only a fraction of the acquisition frame Figure 3.25. We also modified a control interface provided in QuickPALM (Henriques et al. 2010) in order to switch quickly between stroboscopic and non-stroboscopic illumination modes.

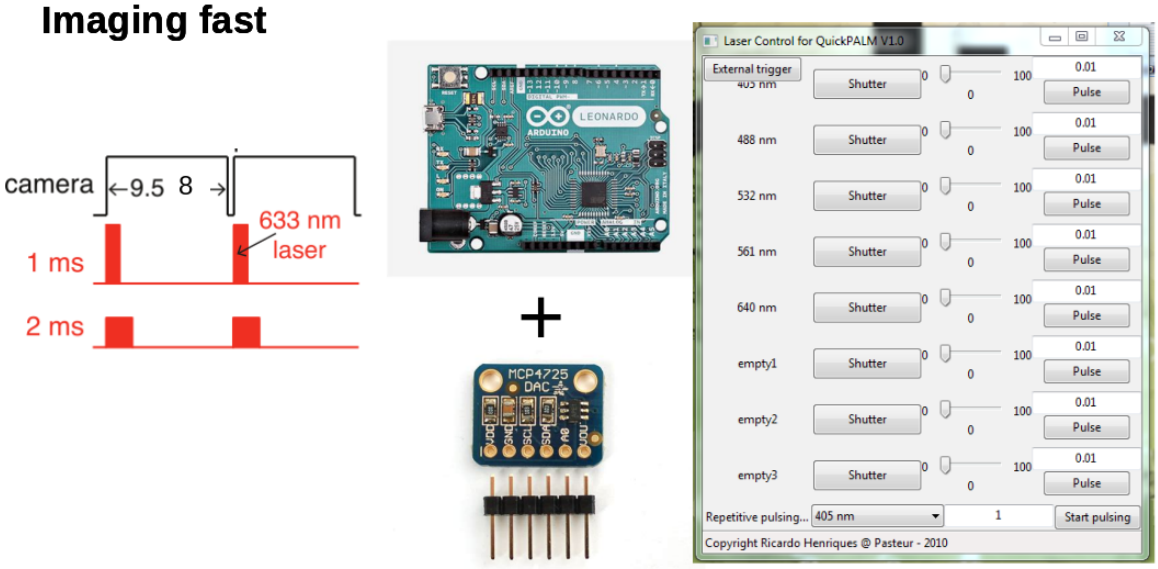


Figure 3.25: **Principle and implementation of a stroboscopic illumination scheme**. (*left*) Principle of stroboscopic illumination. (*middle*) Arduino and Digital-to-Analog converter used to implement stroboscopic illumination. (*right*) Interface (adapted from Henriques et al. 2010) to control stroboscopic illumination.

2. **Cell labeling.** To perform spaSPT (Anders S. Hansen, Woringer, et al. 2018), one has to use photo-activatable fluorophores, such as the ones of the Janelia Fluor series (PA-JF), such as PA-JF$_{646}$ and PA-JF$_{549}$. We thus imaged the c-Myc-Halo cell lines under various conditions of labeling of PA-JF$_{646}$ and PA-JF$_{549}$. Unfortunately, the SPT data obtained was highly inconsistent with previously published data, yielded very few detections, with most of them not being located in the nucleus (Figure 3.26c, compare with Figure 3.26b).

   These results were extremely puzzling because both IF (section 2) and labeling with non-photoactivatable dyes such as JF$_{635}$ or TMR yielded normal nuclear localization. In order to disentangle this issue, we performed imaging with a dual labeling ($\sim$ 50 nM TMR and $\sim$ 100 nM PA-JF$_{646}$). The TMR signal allows to locate the cells, but also to monitor the degradation state of the Myc protein (estimated to be $\sim$ 30 min based on live cell imaging). We acquired movies at 8ms/frame with no stroboscopic illumination. The TMR signal is mostly nuclear (but clearly not exclusively, and this is expected), but most of the PA-JF$_{646}$ detections are perinuclear.

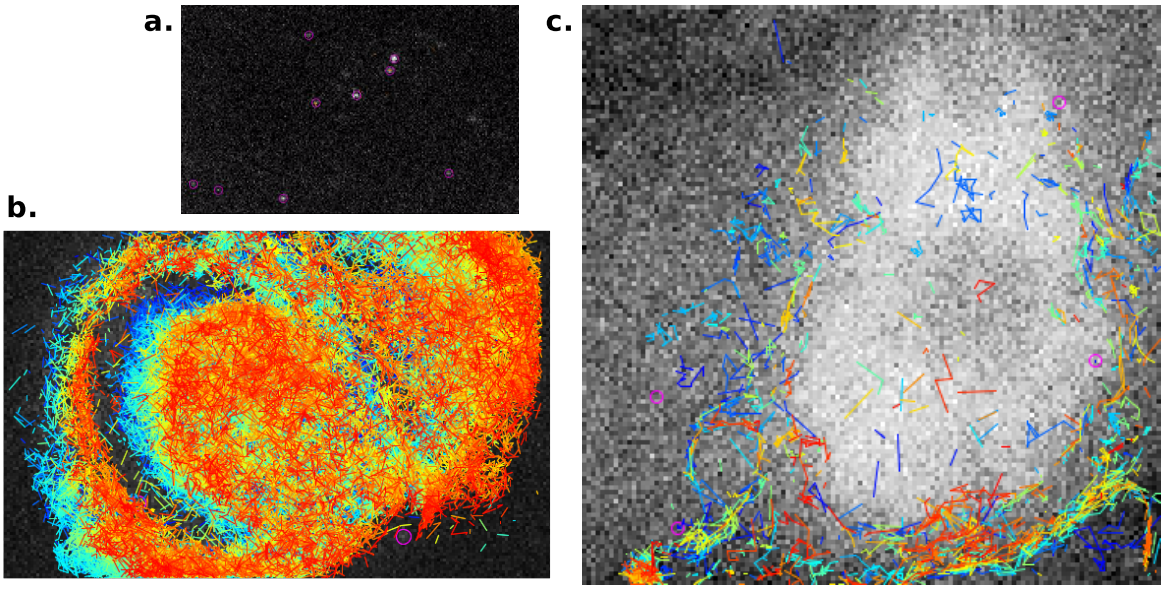


Figure 3.26: **(a)** Sample single-molecule image of c-Myc-Halo-PAJF$_{549}$. The detected molecules are circled in pink. **(b)** Trajectories reconstructed in this experiment, performed in Berkeley. Most of the trajectories are located within the nucleus, as expected, unlike in **(c)** c-Myc-Halo cell dual-labeled with TMR (widefield) and PA-JF$_{646}$ (SPT). Both are conjugated to HaloTags, but display distinct localization patterns, suggesting a fluorophore defect.

Various personal communications (From the Darzacq lab: Anders S. Hansen, Wulan Deng, David McSwiggen, Alec Heckert, the Heliot lab, the Mazza lab and the Dahan lab and directly from Luke Lavis, who synthesized the dyes) sketched a coherent picture in which the PA-JF dyes, unlike the normal JF dyes, tend to degrade over time, no matter how they are stored or handled. Based on this feedback, we decided to abandon the use of PA-JF dyes, and to use non photoactivatable dyes. Unfortunately, due to the time spent troubleshooting these issues, we did not have the time to start optimizing different imaging conditions, in which the density of fluorophores cannot be precisely controlled.

3.5 Discussion and conclusion

In this chapter, we proposed an approach to specifically characterize the sequence-determinants of c-Myc dynamics. As explained in section 3.2, the dynamics of a TF and transcription regulation are deeply intertwined, and a more precise understanding of the domains and interactions determining these interactions are of high basic and applied interest.

In this work, we aimed at characterizing the dynamics of c-Myc in mESC using SPT, and to compare it with various mutants (deletion of some protein domains) and drugs that disrupt c-Myc interactome, with the aim of determining what are the key drivers of c-Myc dynamics.

This discussion is organized as follows: we first review the work we achieved, and provide a critical reading of the difficulties we faced. After a series of perspectives to resolve these issues, we finally envision the experimental perspectives of this work. A more general discussion is presented in section III.

3.5.1 Summary of the work

1. **Generation of a cell line to study c-Myc dynamics.** During the course of the PhD, we developed and validated a biological mESC model to study the dynamics of the c-Myc oncoprotein in live cells. We generated a series of c-Myc knock-in cell lines with various tags (Halo, Dendra2,

mEOS) and validated them using morphological, functional and genomic characterization. The current state of the results strongly suggests that the generated cell lines behave like mESC, and that neither the expression or regulation of c-Myc or key pluripotency genes is dramatically affected, despite statistically significant differences in the expression levels of some pluripotency factors (Oct4, Sox2, Nanog).

Nonetheless, we point out that a definitive validation of whether c-Myc-Halo protein is functional is incomplete/indirect. Our current arguments mostly rely on the fact that c-Myc is a key pluripotency factor and that mESC are expected to show a strong phenotype in the absence of the c-Myc protein (they are expected to display a differentiated phenotype, and severe transcription defects). The absence of such phenotype suggests that the tagged c-Myc protein is functional. Clearly, a definitive demonstration will involve a ChIP-seq experiment, as described in section 5. We performed the first steps of the ChIP-seq protocol and stopped before the library preparation, since we had to re-optimize the imaging protocol.

2. **Determination of key experimental parameters for SPT.** We successfully determined the lifetime of the edited protein and showed that it was similar to the WT cell line, and that c-Myc has a short half-life in mESC (around 20-40 min), as previously reported. We also used a recently published, FACS-based approach to determine the absolute number of c-Myc proteins in a mESC and found it to be around 60000 copies. This number is key to biophysical modeling. Finally, we examined the localization of the fluorescence as the c-Myc protein gets degraded. From these experiments, we deduced that the imaging should be performed promptly after labeling, that is in a window of 5-30 minutes after labeling, in order to mostly observe undegraded proteins.

3. **Preliminary imaging.** We then performed preliminary imaging in fast SPT. The experiment was initially performed in the Darzacq lab, and then we developed a system in Paris to perform similar experiments. The first experiments yielded encouraging results, but recurrent issues with the fluorophores in use lead us to temporarily stop the experiments.

4. **Work in progress.** Indeed, it became clear that the photo-activatable dyes synthetized in Janelia (the PA-JF_{549} and PA-JF_{646} dyes) suffered irreversible degradation, and that they could not be used for SPT. In such a context, spaSPT, as presented in section 1.3 cannot be performed, and one needs to rely on non-photoactivatable fluorophores such as TMR or the regular JF dyes (JF_{549}, JF_{646} and JF_{635} for instance). We are currently exploring this option with the help of Mickaël.

3.5.2 Issues encountered during the project

1. **Choice of serum.** We begin by briefly mentioning the fact that c-Myc is involved in the signalling pathway that maintains cells in an ES phenotype (Cartwright 2005). As such, the choice of the serum is critical, and it is difficult to know which serum conditions should be used. In particular, knowing whether synthetic serum replacements (Tanimoto et al. 2008; Kalaskar and Lauderdale 2014) would provide more reliable results is not known.

2. **Issues with imaging.** Most of the project on c-Myc dynamics relied on stroboscopic, photo-activatable SPT (spaSPT; section 1.3). We generated three different cell lines amenable to spaSPT (one with a HaloTag, one with mEOS and one with Dendra2). Unfortunately, the mEOS and Dendra2 displayed an abnormal phenotype and could not be used for further experiments, and we were left with only one cell line and no option to cross-validate our results with alternative tags. We briefly discuss our difficulties with tagging the C-ter part of c-Myc in section 3. To perform spaSPT with a c-Myc-HaloTag cell line, one needs to conjugate a photoactivatable dye. The only available photoactivatable dyes are the PA-JF_{549} and PA-JF_{646}. Due to stability issues, it became unfortunately impossible to use those dyes for SPT.

Fortunately, it is possible to perform SPT without such dyes, and relies on non-photoactivatable fluorophores. The loss of photo-activation is always performed at the expense of the quality of tracking and of the number of trajectories per cell obtained.

Another limit should be mentioned: the fact that we used only one type of tag, the HaloTag, to perform experiments. In this context, it is very difficult to cross-validate the results we obtained. This cross-validation is important because the HaloTag, like any tag, is known to influence diffusion. It is suspected that the charges of the protein yield some residual non-specific interactions with DNA (L. Xiang et al. 2019).

3. **Issues with the generation of the cell line.** Many reasons can explain why it was technically difficult to obtain homozygous clones of genome-edited c-Myc-Halo (we had to change linker), and that we obtained only a handful clones of c-Myc-Dendra2 clones and c-Myc-mEOS. Furthermore, the success rate of the genome-editing procedure seems to correlate with the choice of the tag.

   This hints to the fact that despite a mild phenotype, the tag on c-Myc is likely to affect the cells in several ways. First of all, since c-Myc is involved in transcription elongation and the tag is located close to the DNA binding domain, it is possible that the binding of c-Myc to DNA is impaired. This issue can be easily assessed, and we initiated a ChIP-seq experiment to confirm it. The results are still pending.

   Second, it is possible that the tag affects other key functions of c-Myc. It is possible that the tag interferes with a known or unknown partner of c-Myc, for instance through steric hindrance with the leucine zipper (LZ) domain, located at the very C-ter of the protein.

   Finally, alternative explanations might relate to the fact that the tag might also alter the dynamic properties of the protein, or its aggregation properties. Further experiments are needed to assess these behaviours.

3.5.3 Experimental perspectives

In addition to the use of non-photo-activatable dyes, our work could be brought forward in several ways.

1. **Automated segmentation.** On a very operational side, many of the quantifications that we performed on images to assess the pluripotency of cells were done manually, which limited strongly the sample size that we could analyze, and forced us to work in 2D. New 3D, deep-learning-based segmentation techniques are being more and more widespread and should be used to perform automated quantification (Qu et al. 2011; Machado, Mercier, and Chiaruttini 2018).

2. **From transfections to induction techniques.** To transfect the c-Myc mutants constructs is a choice mainly driven by simplicity. Indeed, transfections have several drawbacks. First, it is extremely difficult to control the expression level of the protein. In our case, we decided to tackle this issue by performing a dual-labeling: a high-density label to assess the expression level of c-Myc and a low-density level to perform SPT. Such a labeling approach was proposed in (McSwiggen et al. 2018). Second, transfections do not provide timely information, and thus prevents to distinguish between direct and indirect effects.

   In a transfection setting, the plasmid is introduced in the cells around 24 hours before imaging. Other techniques have been pioneered that allowed timely induction of a TF. One seminal approach consisted in the fusion of the c-Myc protein and an estrogen receptor (ER): upon induction of an estrogen analog, the c-Myc-ER fusion translocates to the nucleus (Medh et al. 2001).

   Since then, several techniques have been proposed to either "anchor away" a protein of interest, to "uncage" a factor, or to actively degrade it (Cambridge et al. 2009; Terai et al. 2011; Strickland et al. 2012; Dagliyan et al. 2013; Konermann et al. 2013; Martell et al. 2016; Wehler et al. 2016). These techniques should definitively be investigated to perform cleaner c-Myc imaging experiments.

3. **Characterize the spatial point patterns of Myc/Myc localization.** Second, our dynamics measurement cannot be envisioned without a spatial perspective on c-Myc organization. It is indeed critical to characterize not only the dynamics of c-Myc, but to correlate it with its spatial distribution and the distribution of c-Myc interaction partners. In particular, it would be extremely interesting to locate the phospho-isoforms of c-Myc in the nucleus, following an existing low resolution study (Myant et al. 2015).

   This is usually achieved through super-resolution microscopy (PALM, STORM, PAINT or expansion microscopy; E. Betzig et al. 2006; Hess, Girirajan, and Mason 2006; Rust, Bates, and Zhuang 2006; Manley et al. 2008; Jungmann et al. 2016; Schnitzbauer, Strauss, et al. 2017; Cahoon et al. 2017; M. Gao et al. 2018; Gambarotto et al. 2018, reviewed in B. Huang, Bates, and Zhuang 2009), good error correction (including sample drift; Theer, Mongis, and Knop 2014; Carlini et al. 2015; R. Han et al. 2015; Culley et al. 2017) and corresponding analysis techniques (Ogata and Katsura 1991; Bolte and Cordelieres 2006; Kiskowski, Hancock, and Kenworthy 2009; Helmuth, Paul, and I. F. Sbalzarini 2010; Baddeley, Cannell, and Soeller 2010; P. Sengupta, Jovanovic-Talisman, Skoko, et al. 2011; Veatch et al. 2012; Shivanandan, Radenovic, and I. F. Sbalzarini 2013; P. Sengupta, Jovanovic-Talisman, and Lippincott-Schwartz 2013; Rollins et al. 2015; Mlodzianoski et al. 2011; Levet et al. 2015; Andronov, Lutz, et al. 2016; Andronov, Orlov, et al. 2016; Pageon et al. 2016; Andronov, Michalon, et al. 2017; Andrews et al. 2017; E. A. K. Cohen, Abraham, and Ober 2017; Schnitzbauer, Y. Wang, et al. 2018; Laine et al. 2018; Pike et al. 2018).

   Albeit seducing, all these techniques require the sample to be fixed. Fixation has been a subject of intense debate, and as of today, no artifact-free fixation method exists. It is widely acknowledge that even after long fixation times (tens of minutes), proteins can still move, and that the ultrastructure of cells can be deeply modified (Schnell et al. 2012; Y. Wang et al. 2014). Alternatives to slow and denaturing fixations come from the field of EM, in which fixation is often performed in two steps: first the sample is rapidly frozen, which stops all cellular processes. In a second step, the frozen water is substituted with a regular fixative agent such as paraformaldehyde. This technique is called high-pressure freezing followed by freeze substitution (HPF-FS; McDonald et al. 2007; Schorb and Briggs 2014; Paez-Segala et al. 2015). So far, super-resolution has been extremely challenging to perform on such samples.

4. **Myc aggregation and phase separation properties.** A complementary approach to the imaging of c-Myc in live cells and fixed cells is the characterization of its aggregation properties. Indeed, c-Myc as a transcription factor is known to interact with hundreds of protein partners, and the purified c-Myc protein (Yeh et al. 2004) has been shown to phase-separate *in vitro* (Boija et al. 2018). The exact determinants of this phase-separation behaviour, its effective existence inside cells and the functional consequences of it are not known and should be investigated, as they appear today at the crossroad between spatial organization of the genome and transcription regulation (Hnisz et al. 2017; Sabari et al. 2018; S. Chong, Dugast-Darzacq, Z. Liu, Dong, G. M. Dailey, et al. 2018).

   Key information on the behaviour of c-Myc in cells could come from non-imaging techniques, including *in-cell* NMR (Serber and Dötsch 2001; Selenko and Wagner 2007; Maldonado, Burz, and Shekhtman 2011; Freedberg and Selenko 2014; Barbieri, Luchinat, and Banci 2016). In live-cell NMR, the NMR relaxation properties of a specifically labeled protein of interest are followed, allowing to derive information such as its disorder behaviour in cells or its existence within a complex. Comparison with *in vitro* NMR allows to draw conclusions about the specific cellular environment (see for instance a study on p53 Borcherds et al. 2014, α-synuclein Theillet et al. 2016 and on ataxin Sicorello et al. 2018). c-Myc, as a relatively high expression protein, is a good candidate for such approach.

   Such approaches should be complemented by *in silico* simulations of the behaviour of the protein, using molecular dynamics (Lindahl, Hess, and van der Spoel 2001) of related methods.

It is traditionally difficult to simulate low complexity and intrinsically disordered proteins, and additional difficulties arise when trying to simulate a protein in which some domains (such as the DNA-binding domain) are folded and others (such as the trans-activation domain) are not. Fortunately, recent progress has been made in the simulation of these proteins (see for instance Vitalis and Rohit V. Pappu 2009; Gibbs and Showalter 2016; Robustelli, Piana, and Shaw 2018) and in the simulation of protein-protein interactions between disordered proteins (Grauffel, Stote, and Dejaegere 2010; Zwier et al. 2016; Saglam and L. Chong 2018).

5. **Towards a good null model of Myc dynamics.** In addition to a precise characterization of c-Myc dynamics in the absence of protein domains, a *null model* of c-Myc dynamics should be progressively built. We call a null model a series of proteins that would decouple the various influences of external factors on c-Myc proteins.

   This model could be deduced from various types of experiments. Firstly, as presented above, c-Myc protein-protein interaction inhibitors are likely to provide instrumental information about how c-Myc dynamics are shaped, potentially allowing to determine which (series of) protein-protein interaction governs the dynamics of c-Myc. Second, c-Myc mimetics, such as synthetic domains mimicking the bHLH domain of c-Myc (Ruiz García et al. 2017) might provide another part of the picture. Finally, the use of targeted point mutations to disrupt specific protein-protein interactions will help to finally get a precise idea of the sequence-to-dynamics map of c-Myc.

Chapter 4

Going further: the complex diffusion analysis challenge

4.1 Context

The complex diffusion analysis challenge is an ongoing project that aims at both benchmarking existing complex-diffusion analysis algorithms and at fostering the development of new analysis techniques. It focuses on nuclear and membrane samples and on the FCS and SPT modalities. It was initiated with Ignacio Izeddin, Cyril Favard and Hugues Berry. The following chapter is inspired from a manuscript in preparation:

Anomalous diffusion in live cells : bridging the gap between experiments and models through collaborative challenges. (Maxime Woringer, Hugues Berry, Dominique Bourgeois, Cyril Favard, and Ignacio Izeddin).

4.2 Introduction

The life of a cell is governed by highly dynamics microscopic processes. Two notable examples are the diffusion of membrane receptors and the kinetics of transcription factors governing the rates of gene expression. Different fluorescence imaging techniques have emerged to study (macro)molecular dynamics. Among them, fluorescence correlation spectroscopy (FCS) and single-particle tracking (SPT) have proven to be instrumental to our understanding of cell dynamics and function.

These techniques have unraveled an unforeseen complexity and diversity of mechanisms of protein diffusion. Many efforts have been devoted to analyze datasets generated by FCS or SPT, ranging from diffusion coefficient estimations to inference approaches. However, choosing the appropriate algorithm can be challenging. Indeed, the richness of experimental data often makes it difficult to determine which are the models to be considered and the relevant biophysical parameters to be estimated.

In such a setting, both the imaging and machine learning communities have often relied on collaborative challenges, where labeled (training) and unlabeled (evaluation) simulated data are provided to competitors all over the world. The challenges foster the development of state-of-the-art analysis algorithms. They provide a unified data benchmark based on biologically-relevant metrics in order to compare the diffusion analysis software available for the community.

In a previous section we briefly reviewed key anomalous diffusion models relevant to cell biology and some of the existing techniques to either infer model parameters or perform model selection. We here stress the importance of designing realistic datasets, closely mimicking the type of data obtained in the field by biologists. We highlight often overlooked limitations in current acquisition methods and emphasize the noise levels and expected biases of these techniques.

With this perspective, we hope to provide the molecular imaging community with a comprehensive set of data and metrics allowing to objectively evaluate existing and new analysis tools, as well as

instigating an open discussion about the limitations and challenges of analyzing and modeling diffusion of molecules in the complex environment of the cell.

4.3 From imaging to analysis: strength and gaps

4.3.1 Why we care about anomalous diffusion in cells

As explained in section I.1, to understand how proteins move in a cell is key to the understanding of many biological mechanisms of paramount importance, ranging from the regulation of transcription, signalling and neuro-degenerative diseases.

For instance, how neurotransmitters diffuse inside and outside cells strongly determines many parameters of neuronal transmission. Similarly, the reaction rates of transcription factors are intrinsically coupled to the way they diffuse in cells. There is then a crucial need to better understand how protein diffuse in cells.

Over the years, researchers have developed various imaging-based techniques to assess diffusion in live cells. Among them, two stand out particularly: single-particle tracking (SPT) and fluorescence correlation spectroscopy (FCS). In SPT, a low number of fluorescently-labeled proteins are visible at a given time, so that their density is low enough to discern single molecules. In this regime, individual proteins can be tracked over time, and the coordinates of the protein can be extracted, yielding trajectories that need to be analyzed further. In FCS, a confocal-inspired beam is focused at one given location in the cell and the fluctuations of light intensity collected are recorded. When a fast diffusing molecule crosses the confocal region, its emitted photon are collected for a very short period of time. Conversely, when a slow-diffusing protein crosses the confocal region, it takes it a lot of time to exit the volume, and photons are collected for a longer time. The averaging over many events and corresponding modeling then allows to recover the existing species in the solution.

Using these imaging modalities, researchers progressively mapped the dynamics of a high variety of proteins. They also noticed that often, proteins and particles diffusing in cells did not exactly follow the laws of Brownian diffusion, but rather displayed anomalous diffusion. This is exemplified in (Figures 4.1). Indeed, in (Izeddin, Récamier, et al. 2014), the authors used SPT to characterize the anomalous diffusion of two key proteins in a mammalian cell: c-Myc and P-TEFb. When they computed their MSD, they found that it did not follow a straight line, but could rather be fitted using a power-law with exponent $\alpha \neq 1$, indicating that more complex processes than pure Brownian diffusion happen in cells. The complexity picture is traditionally further reinforced by the fact that it is not only the size of a particle that determines its diffusion properties, but also its chemical properties. This is exemplified in (Etoc et al. 2018), in which the authors used nanoparticles of constant diameter (25 nm) but bearing different chemical reactivities (through different passivations). Surprisingly, despite the fact that the particles had the same size, they exhibited dramatically different diffusive properties, from fast and almost free-diffusing to slow and highly anomalous. These experiments suggest that diffusion in cells is influenced by complex mechanisms, and that elucidating these processes would yield to important discoveries in how a cell is organized.

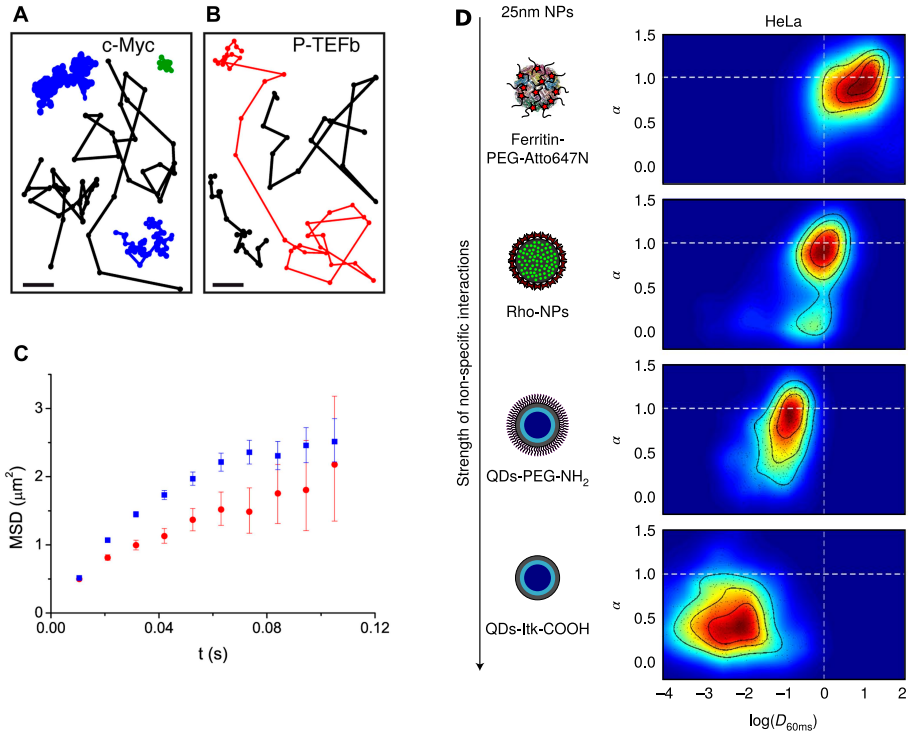


Figure 4.1: **Two examples of non-Brownian diffusion in cells (A,B,C)** anomalous diffusion of the P-TEFb complex. **(D)** Anomalous diffusion of injected nanoparticles in cells. The panel described the distribution of diffusion coefficients (D) and anomalous diffusion exponents (α) for various particles of the exact same size. Source: (A,B,C: Izeddin, Récamier, et al. 2014, D: Etoc et al. 2018)

From a biophysical standpoint, anomalous diffusion can arise from two (non-exclusive) types of mechanisms:

- Because the protein diffuses in a complex, tortuous media. In that case, macroscopic anomalous diffusion comes from the fact that diffusion is *spatially impaired*. Thus, a fine characterization gives information about the space in which the protein diffuses
- Because the protein interacts with other elements of the media, and thus has a tendency to transiently "stick" to it, and to be funneled, depending on the complementarity of the surface of the protein and the structure. Diffusion is *temporally impaired*, as a first approximation.

Crucially, protein dynamics thus not only provide us with key parameters to explain their reactivity, they also carry information about the global cellular organization and the reactivity of the abundant surfaces. From an inverse problem perspective, one can then ask whether one could use SPT, FCS, or a combination thereof to solve the structure and reactivity of a live cell Figure 4.2.

As usual with inverse problems, to recover the structure and reactivity of the space is usually an ill-posed problem. For the case of SPT and FCS, the amount of available data is extremely low compared to the size of the space to be mapped, and it is thus necessary to rely on simplifying hypotheses. These hypotheses are physical or mathematical models of anomalous diffusion of proteins (Figure 4.2). In this setting, it becomes possible to get estimates of the parameters of the models, and to perform model selection, which is a first step to the mapping of the environment of a cell. In the next sections, we discuss the current strengths and limitations of SPT and FCS, and how additional information could be extracted from this data using advanced analysis techniques.

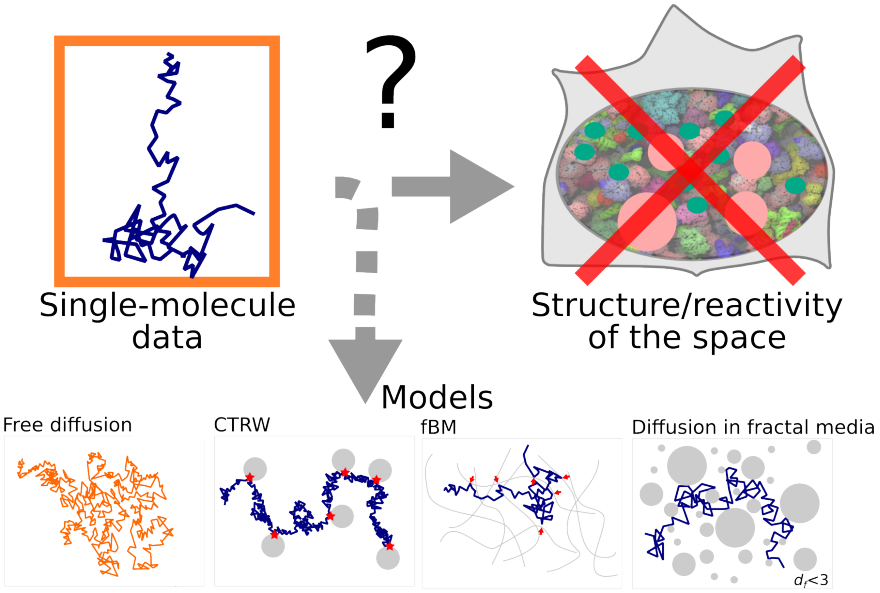


Figure 4.2: Solving the inverse problem of SPT/FCS: to recover the structure and reactivity of the cellular space from SPT or FCS is not possible. Conversely, mathematical models can be used to constrain this inference, and one revert to traditional model selection and parameter inference problems.

4.3.2 State of the art of current data

1. **Membrane tracking traditionally provides high quality data.** For many years, tracking of proteins in membranes provide a gold standard in terms of the quality of the data that can be obtained. Moreover, this field has concentrated many advances that significantly improved data quality. A few of them can be listed:

   Imaging system. the use of Total Internal Reflection Fluorescence (TIRF; Ambrose 1956) provided a very high sectioning (the collected fluorescence comes from $\sim$ 200 nm above the coverslip), and the apparition of PSF-shaping allows to perform 3D SPT.

   Labeling systems. brighter and brighter dyes were designed, initially starting from antibodies conjugated to very stable fluorophores such as Alexa 647, and then to Quantum Dots or nanodiamonds conjugated using nanobodies. These latter allow almost infinite imaging time, since they photobleach extremely slowly.

   In this setting, a membrane SPT experiment traditionally yields thousands of tracks, in which the tracks are long (several tens of points) and with a high pointing accuracy (often below 10 nm). This setting can be regarded as the best imaging conditions that can be obtained in cells.

2. **Current limitations of nuclear imaging.** These ideal membrane conditions have never been reproduced in the cytoplasm or in the nucleus. Several factors explain why imaging inside the cell is more challenging:

   Distance to the coverslip. Reaching the nucleus requires to image at least 1 µm from the coverslip. At this distance, TIRF imaging cannot be applied and more out-of-focus light is collected, leading to reduced signal-to-noise ratio. Several approaches have been proposed to increase the z-sectioning in the nucleus, including Highly Inclined and Laminated Optical sheet (HILO; Tokunaga, Imamoto, and Sakata-Sogawa 2008) or various designs of light-sheet imaging, such as the lattice light sheet (B.-C. Chen et al. 2014). These techniques dramatically reduce the collection of out-of-focus light, but not to the point of TIRF microscopy.

   3D diffusion in cells. Second, unlike in membranes in which proteins diffuse slowly in two dimensions, proteins in the nucleus diffuse in 3D 10-100 times faster. Fast-diffusing proteins

are more difficult to track, require higher laser powers and faster imaging. Moreover, proteins can move in 3D and diffuse outside the focal plane, thus limiting the length of the trajectories obtained by SPT.

Cell permeability. Third, labeling such as quantum dots (QD) or nanobodies are not really available to track proteins in cells. QD are usually too big to freely move in the nucleus, and antibodies/nanobodies are tedious to transfect in cells. Finally, most photostable organic dyes such as Alexa647 or Atto647 are not membrane permeable and thus cannot be used with live cells.

When moving to the analysis step, nuclear SPT is currently often corrupted by the following limitations: high noise levels, short trajectories, poor spatial and temporal sampling. To tackle this issues, two options are available: (1) improve the quality of the data generated (we presented an approach in section 1.2, but see also MINFLUX; Balzarotti et al. 2017) or (2) improve the quality of the analysis.

4.3.3 Limitations of current analysis techniques

Despite the high number of analysis techniques to analyze FCS and SPT (briefly reviewed in section 4.1), several questions remain unanswered and key features missing:

1. **Incorporating good models.** The diversity of anomalous diffusion models is extremely high (see for instance Metzler, Jeon, et al. 2014), but for most of them, it is not known whether they could be relevant to interpret SPT in cells. Furthermore, most of the existing SPT analysis softwares only consider pure diffusion, or mixtures of pure diffusion. Thus, modeling of dynamics data that goes beyond free diffusion is currently a blind spot in the literature.

2. **Adapted to real data.** As briefly detailed in the previous section, nuclear imaging suffers difficult imaging conditions. Many of these conditions have been documented, but most analysis algorithms available on the market are not suitable for real-life datasets. The situation is further complicated by the fact that at the time of writing, virtually no nuclear SPT dataset is available online, seriously impairing the development of new analysis algorithms.

3. **Benchmarked.** Finally, even when algorithms to analyze the available data exist, there is no clear benchmark that allows to opt for one technique or the other.

   To tackle these issues, we took two approaches: (1) In collaboration with Anders Hansen, we proposed a standard to share SPT data, and pushed for the integration of SPT data into the 4D nucleome data sharing platform (section 4.4). (2) In collaboration with Hugues Berry, Cyril Favard and Ignacio Izeddin, we currently organize a collaborative challenge to foster the development of limitations-aware tools (section 4.5).

4.4 Data sharing project

After a rapid inspection of the literature, it appeared to Anders and I that virtually no SPT dataset tracking proteins in the nucleus of live cells was publicly available on the internet. As a first response, we initially shared the SPT data produced in our lab on Zenodo, a general server to share experimental datasets. We then teamed-up with the 4D-nucleome in order to propose a SPT file standard and a structured data sharing platform, in order to make these datasets widely available.

4.4.1 Sharing with Zenodo

We first decided to share the datasets generated for Spot-On (Anders S. Hansen, Woringer, et al. 2018) under a permissive license. The tracked datasets were deposited on Zenodo (1064 movies;

zenodo:834781; doi: 10.5281/zenodo.834781). This initiative was continued, and additional datasets have also been published:

- McSwiggen et al. 2018: (zenodo:1313872; doi: zenodo.1313872)
- Anders S. Hansen, Amitai, et al. 2018 (zenodo:2208323; doi: zenodo.2208323)
- Oomen et al. 2019 (zenodo:1306976; doi: zenodo.1306976)

In total, more than 3000 single-cell movies, and tens of millions of detections are now available online.

4.4.2 The 4D nucleome

The 4D nucleome (4DN) is a NIH-funded consortium that aims at fostering collaboration among the participating labs, in a similar way as the human genome project or the ENCODE project. It is described as : "The 4D Nucleome program aims to understand the principles behind the three-dimensional organization of the nucleus in space and time (the 4th dimension), the role nuclear organization plays in gene expression and cellular function, and how changes in the nuclear organization affect normal development as well as various diseases." (source: NIH website)

It encompasses labs from various countries, mostly focused around two axes: omics (Hi-C and derivative, genomics) and imaging (multiplexed FISH and SPT). The 4D nucleome has a strong open access policy (preprint policy, data sharing policy) and a mature data sharing portal: `https://data.4dnucleome.org/`.

Since the Tjian/Darzacq lab is part of this initiative, we decided to get in touch with other 4DN and non-4DN labs producing SPT data and the 4DN "operational hub" in order to design a SPT data format that could be used to share these datasets, and to develop a suitable extension to the 4DN data platform.

4.4.3 The SPT format

Despite SPT being an old technique, no effort has been devoted into standardizing the data format to share the trajectory data, unlike what is usually seen in the genomics community, were several standards are usually broadly discussed and implemented. To our knowledge, only one article describes a SPT format (Greenfeld et al. 2015), but the project seems to be abandoned (no update on the Github repository since 2014). In 2017, Rigano and Strambio De Castillia 2017 proposed a data structure and a proposed standard on FAIRshare to share SPT data. We got in touch with the authors (the Strambio lab) in order to foster collaboration. Some of the paragraphs in this section are excerpted from the description of the SPT format.

After discussing with the 4DN operational hub (Burak Alver, Koray Kirli), the David Grunwald lab, the Strambio lab (Caterina Strambio, Alessandro Rigano), the Liphardt lab and Joan Ritland, we came with a plain-text, tab-separated format. We then gathered feedback from a wide sample of the nuclear SPT community:

- Tjian-Darzacq lab at UC Berkeley and Sheila Teves
- Xiaojun Ren and lab at University of Colorado, Denver
- Stephan Uphoff at Oxford University
- Brian English and Zhe Liu at Janelia Farm
- Ignacio Izeddin, ESPCI Paris
- Gordon Hager, Diego Presman, David Garcia, Gregory Fettweis, Ville Paakinaho, NIH/NCI

- Robert Coleman, Charles Kenworthy, Albert Einstein College of Medicine
- Caterina Strambio De Castillia and Alex Rigano

Briefly, the SPT format is split in several parts:

1. a metadata part, that describes the biological experiment, the tested conditions, etc. that is handled according to a pre-existing 4DN-data portal standard.
2. A file header, that described the format version, the units used, the available columns (some of them are optional), and the reference to the definition of the localization error provided.
3. The data itself, in a tab-separated format.

4.4.4 The 4D-data platform

Once a SPT format defined among most of the labs operating nuclear SPT, the 4DN operational hub kindly agreed to implement a new data type in the 4DN data portal. This new feature allows to upload, share and browse SPT datasets in a similar manner as the previously available genomics datasets (Figure 4.3).

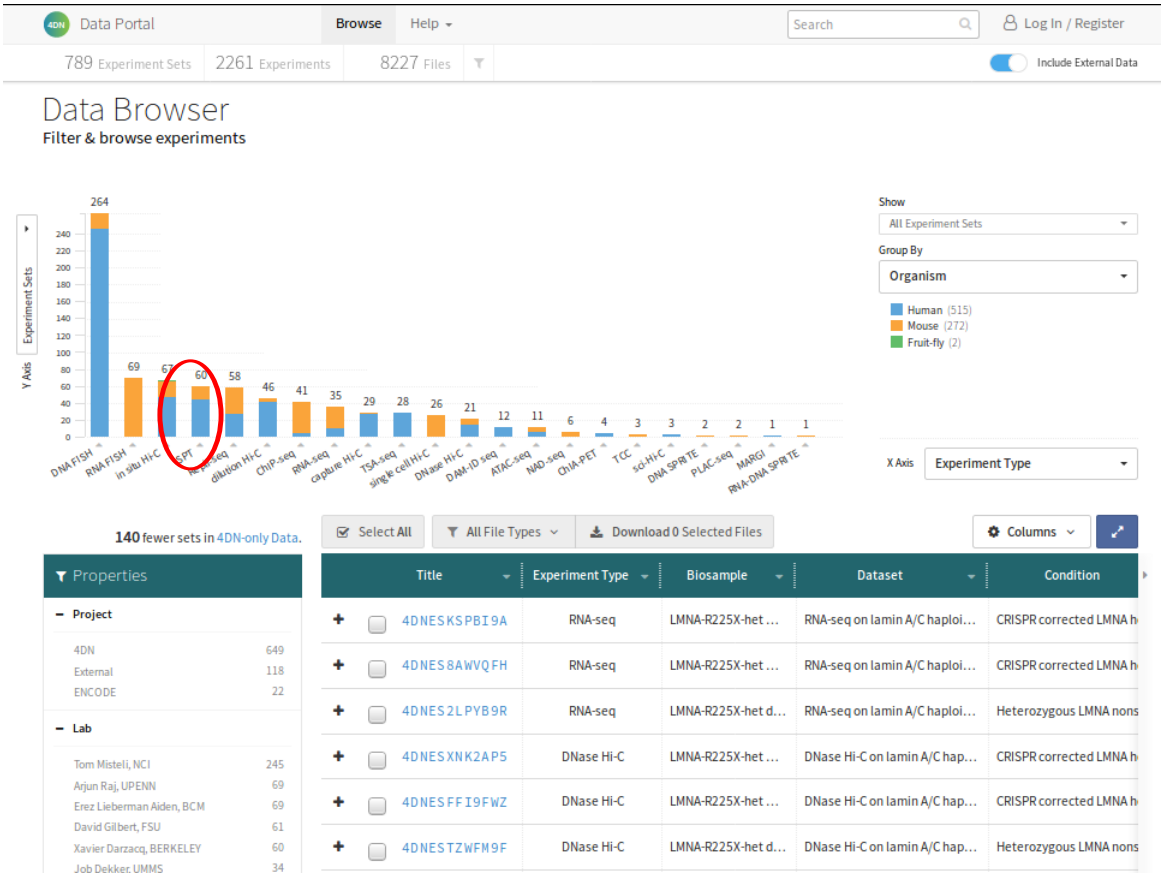


Figure 4.3: The 4DN data sharing platform, with SPT datasets highlighted.

4.5 The complex diffusion analysis challenge

4.5.1 Principle

Since the offer of dynamics analysis algorithms is relatively limited, poorly adapted to currently available data and can rarely compared one with another, we decided to start an international collaborative challenge in order to foster the comparison of existing algorithms and the development of new softwares.

We drew inspiration from challenges in the machine learning community and decided to release two datasets: (1) a training dataset, in which the models generating the data are known, and that can either be used to train a machine learning algorithm or to benchmark it, (2) a test dataset, for which the generating model is kept secret and not known by the participants, allowing for an unbiased benchmark of the tested softwares. This challenge can be seen as a follow-up of the 2D and 3D single-molecule detection challenge (Sage, Kirshner, et al. 2015; Sage, Pham, et al. 2018), and of the tracking challenge (Chenouard et al. 2014).

4.5.2 Modalities

More precisely, we agreed on generating both SPT and FCS data in different modalities (2D in membranes and 3D in the nucleus; Figure 4.4) using `simSPT` (section 1.1). In this setting, we decided to simulate various types of normal and anomalous diffusion (Brownian motion, fractional Brownian motion, CTRW and diffusion in fractals), and mixtures of them (Figure 4.5).

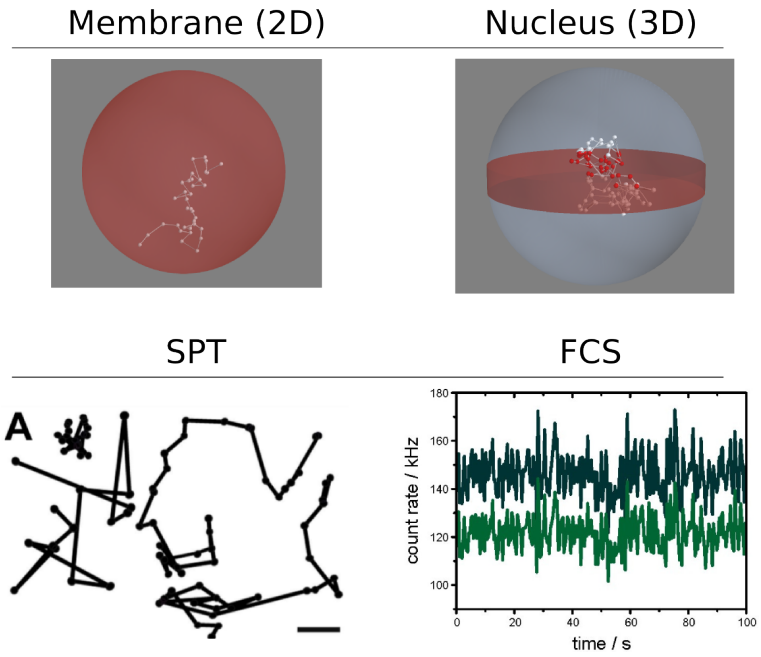


Figure 4.4: Types of imaging modalities considered for the challenge: 2D and 3D setting, and SPT and FCS.

The participants can then get involved in two classes of challenges: (1) a *parameter inference* challenge, in which the generating model is known, but not the exact value of the parameters, and (2) a *model selection* challenge in which both the right model and the right parameters have to be inferred.

The challenge was announced at a few national conferences and a mailing list has been created to facilitate the discussion around the data generation procedure (diffusion.challenge@services.cnrs.fr).

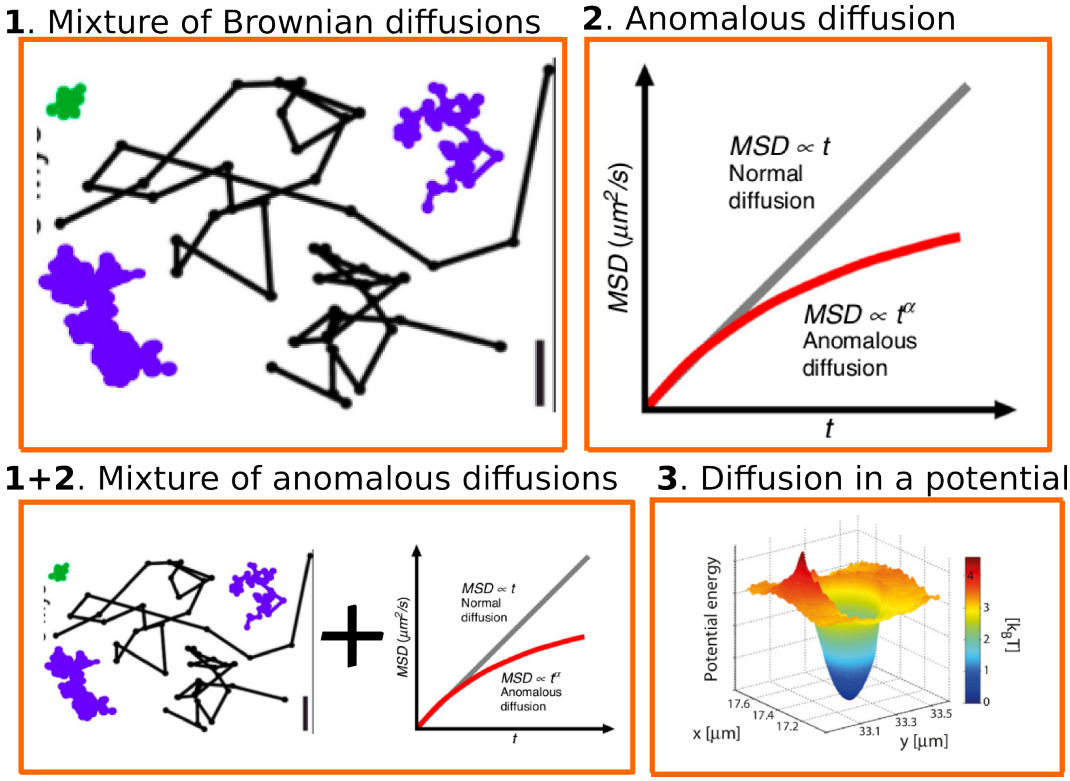


Figure 4.5: Types of models simulated for the challenge.

4.6 Conclusion

Unlike the genomics community, the imaging community really lacks structure and clear objectives. For instance, little to no infrastructure exists to store imaging data, most raw images are not publicly available, and it is often the case that PIs will refuse to share their raw imaging data. Furthermore, despite the existence of image formats and standards, there is neither an accepted standard to share processed data (such as the BED files in genomics for instance, to share ChIP-seq peaks), or clear consensus on which questions the analysis techniques should answer, not to mention a consensus on a way of analyzing data.

Structuring the image community is a long process, that has started on many fronts. Here, we participated in two manners: (1) by the development of a simple SPT standard, and by creating a SPT entry on the 4DN data portal, in order to host the SPT datasets generated among the labs belonging to the consortium. This initiative is followed in parallel by a project to put all FISH imaging data online. (2) By the elaboration of a collaborative challenge to analyze single molecule data.

To be honest, the goal of the challenge is not really to show which software is the best, but rather to foster common thoughts on what should be the metrics to choose one analysis algorithm or another, what are the main caveats of the existing experimental setups, and what are the limits on our ability to analyze noisy data. Thus, we hope to create a community around the challenge, in which SPT and FCS analysis techniques could be openly discussed.

As many ongoing projects, the future will tell whether this initiative is successful.

Part III

Discussion

Chapter 1

Wanted: the map of the nucleus

1.1 Introduction

In this section, we briefly summarize our work. Rather than listing many experimental discussions, we refer the reader to the cited sections for specific discussions. We present in this section some more general discussions and perspectives. We decided to focus on the quest of the "map of the nucleus".

1.2 Summary of the work

First, in the introduction, we identified the strength and challenges faced by researchers in the field of transcription regulation (section 3.1). One of them is the need for information about the dynamics. SPT appeared as the technique of choice to investigate TF dynamics.

1.2.1 Towards tools to analyze SPT

1. **Spot-On.** After identifying the limitations of existing SPT analysis techniques, we designed Spot-On (section 1.3; Anders S. Hansen, Woringer, et al. 2018). Spot-On addresses several limitations of conventional SPT. It can work even with a large collection of very short trajectories, it accounts for defocalization bias and robustly infers existing sub-populations and localization error. We extensively benchmarked Spot-On in many conditions, and evaluated the sensibility to many parameters. In a second part of the PhD, we improved Spot-On to take anomalous diffusion into account (section IV.2). We also theoretically derived the distribution of angles between consecutive jumps under a Brownian motion with noise hypothesis (section IV.3). This opens new avenues to analyze protein motion in the nucleus.

   Together with the Spot-On software (Anders S. Hansen, Woringer, et al. 2018), we proposed stroboscopic, photo-activatable SPT as a reliable approach to study protein dynamics in the nucleus. This technique relies on the use of HaloTag and photo-activatable dyes. These fluorophores allow to acquire a high number of high quality trajectories per cell, and open the way to the characterization of single-cell dynamics. However, the degradation of the PA-JF dyes question the strategy adopted in the manuscript, since no other synthetic photo-activatable is available on the market. Thus, spaSPT cannot be performed anymore, until new synthetic photo-activatable, membrane permeable fluorophores are synthesized or more photostable photo-convertible fluorescent proteins are discovered. This is sad.

2. **simSPT.** In order to benchmark Spot-On, we developed a simulation tool. This tool was designed to simulate SPT trajectories and incorporates biases and limitations usually encountered when imaging mammalian nuclei, including the limited axial detection range, the high localization error and the effect of confinement. We then extended the tool to simulate non-Brownian diffusion.

3. **Complex diffusion analysis challenge.** The tools we developed will not be the ultimate solution to the current limitations of SPT analysis. Rather, a culture of openness and collaboration between experimentalists and theoreticians is needed to address the issues raised by SPT. Building on the fact that little to no SPT of proteins in the nucleus is available, we developed, in collaboration with the 4D-nucleome consortium (1) a documented SPT format, so that SPT trajectories can be shared without ambiguities, (2) we worked with the operational hub of the 4D-nucleome consortium to put the SPT data generated in the lab online and (3) together with Cyril Favard, Hugues Berry and Ignacio Izeddin, we started the *complex diffusion analysis challenge* in order to foster collaboration around analysis techniques, to identify blocking issues and encourage the development of new analysis methods.

4. ***In vitro* imaging.** Once these analysis techniques were developed, we applied Spot-On to a setup where most existing tools available will perform poorly: the case of extremely fast *in vitro* diffusion. In such a setting, most labeled proteins move out-of-focus in less than two frames, and long trajectories are extremely rare. Because Spot-On can build displacement histograms from single jumps, it can analyze this type of data without major issues. In collaboration with Zhijie Chen and Alan Shaw, we used Spot-On to demonstrate that the presence of the substrate of an enzyme did not increase significantly its diffusion coefficient. These experiments bring valuable data in a longstanding controversy (detailed in section II.2).

1.2.2 A model to study c-Myc diffusion

Once the analysis tools presented above were validated, it became possible to apply them to a transcriptionally-relevant problem. We focused on elucidating the sequence-determinants of c-Myc diffusion, namely, identifying the protein domains that drive c-Myc diffusion. An answer to this question would advance our basic understanding of transcription factors biology in general, and of c-Myc in particular. For instance, the determinants of c-Myc *in vivo* sequence specificity are not known, and rely on domains located far from the DNA binding domain (J. Guo et al. 2014; W. F. Lim et al. 2016).

During the course of this PhD, we developed and successfully validated a set of tools to answer this question. We developed a mESC system in which we used the CRISPR-Cas9 system to homozygously knock-in a HaloTag in C-ter of the c-Myc protein. The HaloTag can then be used for SPT imaging. We validated that the cell line displays ESC morphology and phenotype.

However, our results are incomplete, for several reasons. First of all, some of the validations we performed are still incomplete. We performed a RNA-seq experiment to check to what extent the transcriptome of the edited cells differs from the non-edited cells. The samples have been sent for sequencing. We also wanted to verify to what extent the c-Myc binding profile on DNA was modified by the addition of the tag. Due to difficulties with imaging, we decided not to send the samples for sequencing so far. Second, we could not perform reliable SPT imaging, due to the issues mentioned above with degraded photo-activatable fluorophores. Troubleshooting the issues with fluorophores took us several months, and we did not have the time during the course of the PhD to fully implement an imaging technique with other fluorophores. Nonetheless, this approach is now in progress, and we hope to be able to get SPT images in the next few months.

1.3 Perspectives on dynamics: from atoms to a full nucleus

1.3.1 Dynamics as the result of protein-protein interactions

Transcription factors that diffuse in the nucleus of a cell is influenced by many factors, ranging from the individual properties of the protein to collective emerging behaviours of collections of macromolecules in the nucleus.

1. **Effect of size.** In the traditional framework of diffusion, the size of the diffusion molecule is thought to be the major determinant of the diffusion coefficient, that is inversely proportional

to the hydrodynamic radius of the protein. This framework, however, is of little use in cells. Indeed, the cytoplasm and nucleoplasm of a cell cannot be characterized by a single viscosity coefficient, and rheology-like experiments have shown that at least a size-dependent viscosity coefficient should be applied. Such a setting often arises when particles diffuse in a complex mesh or within a crowded medium.

2. **Charges patterns and diffusion.** Second, diffusion not only depends on the physical parameters of the diffusing medium and the proteins, but only on its chemical environment. This has been shown repeatedly. For instance, (Etoc et al. 2018) looked at the diffusion behaviour of 25 nm diameter nanoparticles and found that depending on the chemical properties of the surface, the particles displayed dramatically different diffusion properties (Ferritin was diffusing relatively fast and showing normal diffusion whereas quantum dots coated with carboxylic acids were diffusing up to 1000 times slower and displayed a diffusion much more anomalous. This qualitative change in the diffusive properties connects with studies on intrinsically disordered proteins and the conditions required for phase separation of some proteins. It has been shown theoretically (Vekilov 2010) and experimentally (Pak et al. 2016) using solutions with compounds of increasing charges, that the aggregation behaviour of proteins was modulated by the charge of the protein and the presence of charges in the environment.

   This is particularly interesting in the setting of transcription factors, that usually bear a disordered trans-activation domain. The chemical properties of this domain have been shown to be relevant for phase-separation (Boija et al. 2018) and for the control of transcription (Sherry et al. 2017). Moreover, many transcription factors are under tight regulation by post-translational modifications (PTMs). For instance, c-Myc can be phosphorylated, acetylated, ubiquitylated, SUMOylated, etc. The addition of these residues have the potential to totally modify the set of nuclear components it interacts with. As such, there is a need to characterize the link between PTMs and diffusive properties of TFs.

3. **Towards a map of protein-protein (weak) interactions.** This link can be investigated in several ways, and we proposed a basic approach in (Woringer and Darzacq 2018). A much clearer picture of the dynamics at play in the nucleus could be derived by a more precise knowledge of the "interaction profile" of a TF with the abundant constituents of the nucleus (proteins, RNAs, DNA, etc). With the knowledge of this map, it would be easier to determine the effect of disrupting a given protein-protein interaction.

   To determine this map of protein-protein interactions, several techniques could be useful. A first one is NMR, that was used to characterize the order/disorder state of proteins *in vivo* (Borcherds et al. 2014; Theillet et al. 2016; Sicorello et al. 2018) and to determine the activity of kinases on various substrates (Thongwichian et al. 2015).

1.3.2 Dynamics of proteins on DNA

In addition to protein-protein interactions, many other interactions deserve some interest: protein-DNA, protein-RNA, RNA-RNA, RNA-DNA, protein-other, etc. We present here some thoughts on protein-DNA interactions, since they are key to transcriptional regulation, but also appear fundamental to the formation of some cellular aggregates (S. Chong, Dugast-Darzacq, Z. Liu, Dong, G. M. Dailey, et al. 2018).

It is usually thought that a TF can interact with DNA through two different modes: a "specific" and a "non-specific" mode. This distinction was initially put forward by the study of the LacI protein (Marklund et al. 2018), in which this prototypical TF adopts structurally distinct conformations when "scanning" DNA and interacting specifically with its cognate DNA binding motif. From this seminal finding, it was logically deduced that specific and non-specific interactions would have different kinetic and diffusive signatures. Several experiments that deleted the DNA-binding domain of a TF indeed

found that its residence time on DNA was affected (see for instance Anders S. Hansen, Pustova, et al. 2017).

However, the picture of a clear structural difference between "specific" and "non-specific" interactions has been challenged. A first series of evidence comes from the study of other TFs, such as HoxD9 (Iwahara, Zweckstetter, and Clore 2006), that adopts very similar conformations when binding specific and non-specific DNA. Second, even in the absence of any DNA-binding domain, the proteins usually retain some affinity for DNA (see for instance Anders S. Hansen, Pustova, et al. 2017). It was hypothesized that the charge pattern, and especially the nuclear localization signal (NLS) could have some residual affinity for DNA. In that case, the affinity for DNA is more the result of the chemical properties of the factor than a regulated parameter of the TF. The distinction between specific and non-specific binding was also discussed in length in (Mueller, Stasevich, et al. 2013). Moreover, proteins can undergo non-first order reactions, and especially have DNA-interaction times that do not follow a single-exponential, but rather a power law. In that case, the concept of specific/non-specific interactions, or even slow/fast-binding should be reconsidered (see Grossman-Haham et al. 2018 for an evidence of power-law kinetics at the single-molecule level). Finally, it was shown in (Raccaud et al. 2018) that properties such as the charge composition of the DNA-binding domain (accounting for non-specific interactions) explained the bound fraction of the TF on DNA better than ChIP-seq binding profile (accounting for specific interactions). All-in-all, these findings seriously question the definition of a "specific" and "non-specific" interaction. Is a "long" interaction characteristic of a TF binding to its cognate site? Or is our model of TF-DNA interaction still incomplete. In any case, there is room for further research.

1.3.3 Large-scale macrodynamics

At a larger scale, proteins can adsorb on many structures, described above as "structures of reduced dimensionality". These structures include several phase-separated organelles, and potentially chromatin itself. Several of them were discovered very early in the history of biology (such as the nucleolus), and others were discovered more recently. These structures are dynamic and appear more and more connected to the rest of the nucleus (see for instance research on the structure and dynamics of the nuclear speckles, that strongly correlate with A compartments and transcriptionally active regions (David L. Spector and Lamond 2011; Fei et al. 2017; W. Chen et al. 2018; J. Kim et al. 2018).

To understand diffusion in a nucleus, a precise characterization of these structures is needed: what are the proteins that diffuse "freely" within and across them, what are the molecules that are trapped, etc. For instance, the phase-separated heterochromatin compartment is regulated by the HP1α protein, and clearly excludes some proteins (A. G. Larson et al. 2017; Strom et al. 2017).

For years, chromatin and chromatin compartments have been seen as the major architectural features of the nucleus, and the main drivers of nuclear organization and rearrangements. It now becomes more and more supported by the data that membrane-less organelles can also play an architectural role, and are crucial in the organization of the nucleus. They sequester molecules, move passively and actively, assemble and disassemble, respond to PTMs cues, etc.

1.3.4 The map of the nucleus

1. **Pending questions in nuclear biology.** Taking one step back, one key element in our understanding of nuclear biology is as simple as a map of the nucleus. Indeed, basic questions such as the following have as of today no clear answer:

    - We know how DNA is arranged in the nucleus, but how is RNA arranged, what are the most abundant RNAs in the nucleus, how to they arrange with respect to DNA?
    - Are there zones of locally high macro-molecule concentration in the nucleus? Where are proteins concentrated? Do these concentrations correspond to a specific membrane-less

organelle ? To a transcription site? Same question for RNAs: do they co-localize with DNA or another nuclear structure?

- What is the dynamic range of protein concentrations in the nucleus? Same question for RNA?

2. **What is a map of the nucleus?** Following these questions, we propose that one the main missing piece to understand nuclear organization of function is a 3D super-resolution map of the nucleus, that would overlay DNA, RNA and proteins. Here the idea is not to get the specific coordinate of each RNA species, protein copy and DNA locus, but rather to know at the scale of ~ 15 nm, how are proteins, DNA and RNA arranged, without specificity.

   Such a map would allow to clarify the following points: (1) the proximity/exclusion relationship of DNA wrt. RNA, (2) how diffusion of a TF can be envisioned in the cell: is it impaired by DNA, RNA, protein complexes, etc. (3) Provide an exhaustive vision on phase-separated compartments.

3. **Existing approaches.** We are not aware of any study that presents such a map. Conversely, some efforts were developed that go in that direction. First of all, the spatial organization of chromatin has been mapped at high resolution using Cryo-EM using the Chrom-EMT technique (Ou et al. 2017), providing a precise map of DNA and nucleosome organization at the single-nucleosome level. Second, sequencing approaches such as (Bell, Jukam, Teran, Risca, O. K. Smith, Johnson, J. M. Skotheim, et al. 2018) have mapped RNA-DNA interactions, at the population level. Third, recent imaging papers imaged both RNA and DNA in super-resolution microscopy and found an anti-correlation between DNA-rich and RNA-rich regions (Hilbert et al. 2017; Miron et al. 2019).

   Taken together, one key element is totally missing from these studies: the spatial distribution of proteins. Several approaches could be undertaken to derive this map. A first one relies on unspecific labeling of proteins followed by super-resolution fluorescence microscopy (see for instance the approach in Rhee et al. 2013, that might be repurposed for fluorescence imaging using streptavidin-labeled proteins as fluorescent probes). Another one relies on a seemingly abandoned technique called electron-spectroscopic imaging (ESI) (Bazett-Jones et al. 2008). In ESI, incident electrons that cross the specimen lose some energy when contacting some elements. This energy loss is element-specific. This property was used to identify in the same cell RNA-rich, DNA-rich and protein-rich regions. However, at that time, the resolution was too low to perform an in-depth analysis. Since 2008, I have not found any development in this field. These two approaches constitute two options to derive this map of the nucleus.

1.3.5 Uncovering the links between the structures in the nucleus

In other terms, many proteins with dramatically different diffusion properties coexist within a cell. Some proteins do not appear to move much, sometimes for unknown reasons (e.g: histones; Figure 1.1a). Others move, but are channeled on a given structure of reduced dimensionality, such as depicted in (e.g: nucleolus-associated proteins; Figure 1.1b). Multiple structures of reduced dimensionality can coexist, and carry different types of proteins (chromatin-associated proteins; Figure 1.1c). Despite the complex motion of some proteins, other proteins seem to diffuse without interacting with any of the existing structures (e.g: c-Myc; Figure 1.1d), and other have uncharacterized behaviours (Figure 1.1e).

It is striking to realize that all these proteins actually live in the same volume, the same complex nuclear space, despite the fact that they seem to experience a dramatically different environment (Figure 1.1f). Single molecule measurements only reveal the dynamics of a protein at a time, without providing information about the surrounding space (Figure 1.1g-h). Only a clear understanding of the relationship between the diffusing molecules and the surrounding environment will allow to explain their observe behaviour.

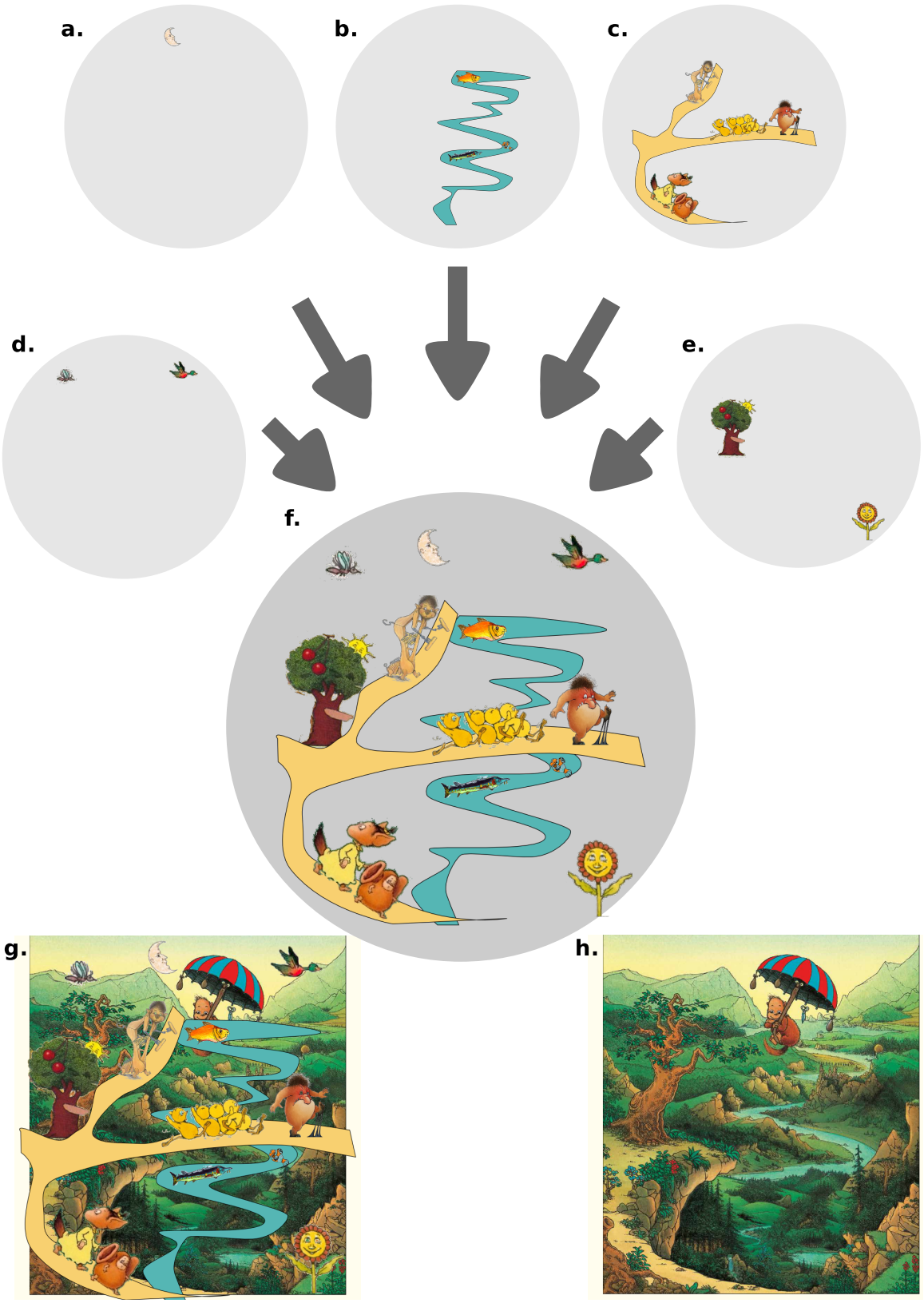


Figure 1.1: **A living forest, a metaphor of the nucleus.** Images from the rich universe of Claude Ponti.

Part IV

Supplementary

Chapter 1

Papers published

In addition to the literature cited in this dissertation (Woringer, Darzacq, and Izeddin 2014; Woringer, Darzacq, Zimmer, et al. 2017; Hansen et al. 2018; Woringer and Darzacq 2018; Parmar, Woringer, and Zimmer 2019), I was involved in a project in the field of epidemiology, that lead to several publications: Mueller et al. 2017; Koutangni et al. 2018; Woringer et al. 2018

In relation with this dissertation

Hansen, Anders S. et al. (2018). "Robust Model-Based Analysis of Single-Particle Tracking Experiments with Spot-On". In: *eLife* 7, e33125. ISSN: 2050-084X. DOI: 10.7554/eLife.33125. URL: https://elifesciences.org/articles/33125 (visited on 01/25/2018).

Parmar, Jyotsana J., Maxime Woringer, and Christophe Zimmer (2019). "How the Genome Folds: The Biophysics of Four-Dimensional Chromatin Organization". In: *Annual Review of Biophysics* 48.1, null. DOI: 10.1146/annurev-biophys-052118-115638. pmid: 30835504. URL: https://doi.org/10.1146/annurev-biophys-052118-115638 (visited on 03/12/2019).

Woringer, Maxime and Xavier Darzacq (2018). "Protein Motion in the Nucleus: From Anomalous Diffusion to Weak Interactions". In: *Biochemical Society Transactions*, BST20170310. ISSN: 0300-5127, 1470-8752. DOI: 10.1042/BST20170310. URL: http://www.biochemsoctrans.org/content/early/2018/07/29/BST20170310 (visited on 08/01/2018).

Woringer, Maxime, Xavier Darzacq, and Ignacio Izeddin (2014). "Geometry of the Nucleus: A Perspective on Gene Expression Regulation". In: *Current Opinion in Chemical Biology* 20. 00022, pp. 112–119. ISSN: 13675931. DOI: 10.1016/j.cbpa.2014.05.009. URL: http://linkinghub.elsevier.com/retrieve/pii/S1367593114000647 (visited on 04/04/2017).

Woringer, Maxime, Xavier Darzacq, Christophe Zimmer, et al. (2017). "Faster and Less Phototoxic 3D Fluorescence Microscopy Using a Versatile Compressed Sensing Scheme". In: *Optics Express* 25.12. 00000, p. 13668. ISSN: 1094-4087. DOI: 10.1364/OE.25.013668. URL: https://www.osapublishing.org/abstract.cfm?URI=oe-25-12-13668 (visited on 06/11/2017).

In the field of epidemiology

Koutangni, Thibaut et al. (2018). "Compartmental Models for Seasonal Hyperendemic Bacterial Meningitis in the African Meningitis Belt". In: *Epidemiology and Infection*.

Mueller, Judith E. et al. (2017). "The Association between Respiratory Tract Infection Incidence and Localised Meningitis Epidemics: An Analysis of High-Resolution Surveillance Data from Burkina Faso". In: *Scientific Reports* 7.1, p. 11570. ISSN: 2045-2322. DOI: 10.1038/s41598-017-11889-4. URL: https://www.nature.com/articles/s41598-017-11889-4 (visited on 01/19/2018).

Woringer, Maxime et al. (2018). "Atmospheric Dust, Early Cases, and Localized Meningitis Epidemics in the African Meningitis Belt: An Analysis Using High Spatial Resolution Data". In: *Environmental Health Perspectives* 126.9, p. 097002. DOI: 10.1289/EHP2752. URL: https://ehp.niehs.nih.gov/doi/10.1289/ehp2752 (visited on 09/07/2018).

Chapter 2

Spot-On v2

The version of Spot-On presented in the (section 1.3) has been validated against a wide range of conditions, both using realistic simulated data and against experimental data, making it a tool that can be used by biologists to obtain trustable results.

Nonetheless, Spot-On still suffers some limitations, both in terms of available features and in terms of choice of the fitting method. More specifically, the limitations and how I tried to overcome them were as follow:

- Choice of the fitting algorithm: the fitting routine used in Spot-On (all versions) is not the right one, from a mathematical standpoint. We thus explored alternative fitting routines (section 2.1).
- Choice of the model: although a multi-state, Brownian model is suitable for the study of many diffusing proteins and protein complexes, most nuclear factors are known to exhibit a variable degree of anomalous diffusion (see for instance Woringer and Darzacq 2018). We thus explored how simple anomalous diffusion models could be implemented in Spot-On (section 2.2).

2.1 Improved fitting

2.1.1 Issues with least-squares fitting

In the published version of Spot-On, the mixture model is fitted to the empirical jump length distribution $\hat{P}(r, \Delta\tau)$ using a non-linear least squares routine. The parameters are optimized within a given range specified by the user. In mathematical terms, Spot-On determines a set of parameters S_{fit} (for instance $S_{fit} = (D_{free}, D_{bound}, p, \sigma)$) by fitting a parametric model $M(S_{fit}, S_{user}, r, \Delta\tau)$, in which S_{user} are user-provided parameters, such as the axial detection range Δz. We denote the estimated parameters by $\hat{S}_{fit}$. Using least-squares, Spot-On computes $\hat{S}_{fit} \in \Sigma_{fit}$ as follows:

$$\hat{S}_{fit} = arg \min_{S \in \Sigma_{fit}} ||\hat{P}(r, \Delta\tau) - M(S_{fit}, S_{user}, r, \Delta\tau)||_2,$$

where $||.||_2$ represents the ℓ_2 norm.

Nonlinear least-square optimization algorithms offer many benefits. The main one is speed. Indeed, significant speed-up are obtained because several second-order methods have been implemented (nickmayorov 2017). The second main advantage is that under a Gaussian model (the errors on the estimate of P are normally distributed), then a least-square formulation of the problem is mathematically equivalent to a maximum likelihood estimate, thus opening the way to derive well founded statistical estimates, errors, etc.

In the case of Spot-On, the non-linear least-squares is performed on the jump-length distribution, and the errors on the estimate of the jump-length distribution are not normally distributed (one cannot write $\hat{P}(r, \Delta\tau) = P(r, \Delta\tau) + \epsilon$, with ϵ normally distributed. As such, it is incorrect to assume that least-squares optimization of the parameters when fitting a theoretical PDF on the empirical distribution of

jump lengths will produce a result equivalent to a maximum likelihood estimate: distributions cannot be compared using a ℓ_2 norm.

One solution to this problem would be to derive a maximum likelihood approach that uses the individual jump lengths (and not its distribution as inputs). This approach seems relatively straightforward but we did not pursue it.

2.1.2 The Kullback-Leibler divergence

Another approach consists into finding a better metric to compare empirical and theoretical distributions. Metrics suitable to compare distributions include the Bhattacharyya distance and the Kullback-Leibler (KL) divergence. We decided to reframe our optimization problem in order to minimize the KL divergence rather than the ℓ_2 norm between the empirical jump length distribution $\hat{P}$ and the model M. Now, $\hat{S}_{fit}^{KL}$ is defined as follows:

$$\hat{S}_{fit}^{KL} = arg \min_{S \in \Sigma_{fit}} D_{KL}\left(\hat{P}(r, \Delta\tau) || M(S_{fit}, S_{user}, r, \Delta\tau)\right),$$

with the KL divergence, $D_{KL}(P||Q)$, between two continuous distributions P and Q (with densities p and q, resp.) defined as:

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \tag{2.1}$$

2.1.3 Comparing distributions using KL divergence

It is usually straightforward to compute D_{KL} between two continuous distribution or between two discrete distributions (using a discretized version of equation 2.1). However, in the Spot-On setting, we derive a discretized version of the jump length distribution (a histogram whose binning is user-specified) and want to minimize the distance with a continuous model M.

Several approaches were develop in order to perform such comparisons. One of the first approach developed relied on performing a kernel density estimate of the discrete distribution to turn it into a continuous one, and then to compute the continuous KL divergence. This approach is usually time consuming. We decided to rely on a different approach described in (Pérez-Cruz 2008), in which a convenient discretization is applied to the continuous distribution, allowing to compute the discrete D_{KL}. An implementation in the `R` language has been proposed.

2.1.4 Optimization

Once a metric, D_{KL} has been chosen, one needs to implement an optimization routine. Indeed, the Levenberg-Marquardt solver used in Spot-On only works to minimize ℓ_2 norms and is unsuitable to KL divergence.

A natural approach to this optimization problem would be to use cross-entropy optimization Botev et al. 2013, a Monte-Carlo, population-based optimization technique based on the KL divergence. Cross-entropy optimization is particularly suited for solving difficult optimization problems, and might be of high relevance when the number of parameters increases (for instance as additional diffusion states are added). However, as most Monte-Carlo, population-based methods, cross-entropy optimization is expected to be relatively slow, even in the most optimized implementations (Benham et al. 2015).

We thus decided to use the standard non-linear optimization routine provided in the Scipy package: scipy.optimize.minimize, that uses the BFGS method (Broyden-Fletcher-Goldfarb-Shanno).

2.1.5 Results

A preliminary version has been released on the Spot-On development website (git repository). Preliminary tests indicate that the method provides valid results, but an in-depth characterization is deeply needed before this KL-based can be used by biologists.

2.2 Anomalous diffusion

Spot-On provides fits of remarkable quality when applied to simulated data (section 1.3) and to *in vitro* diffusion data (figure II.2). However, when moving inside living cells, one usually sees bigger deviations from the theoretical model. Indeed, jump length displacement histograms of proteins diffusing inside the nucleus are often imperfectly fitted using a 2-state model, because the immobile fraction seems to expand a little bit. In the current version of Spot-On, this is usually solved by adding a third diffusion state to the model, and fitting a 3-state model with two free states.

One can question the relevance of this approach, however. As of today, it is to our knowledge unknown whether a simple non-Brownian diffusion model can fit the available data.

We thus undertook the following approach, in which we decided to first use `simSPT` to simulate anomalous diffusion. In parallel, we performed a literature search in order to obtain the propagators (theoretical jump length displacement distributions) for various anomalous diffusion models.

The project was motivated by Anders Hansen, and the literature search was performed jointly. We present preliminary results in this section.

2.2.1 Simulation of anomalous diffusion using `simSPT`

We used `simSPT` (section 1.1) to simulate anomalous diffusion models and computed their empirical jump length distribution using Spot-On, for various levels of anomalous diffusion. Figure 2.1 presents how the simulated jump length distribution evolves under fractional Brownian motion as an example.

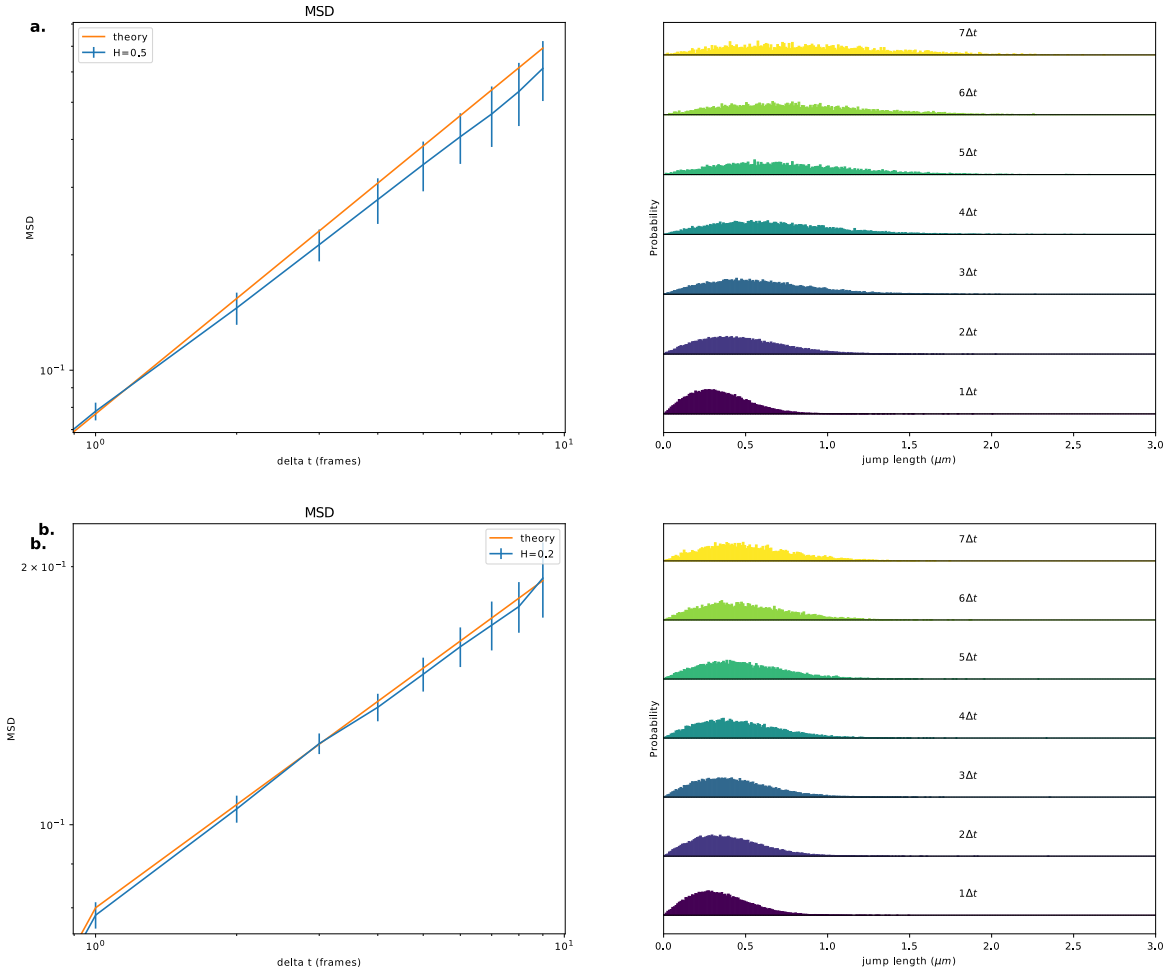


Figure 2.1: MSD (*left*) and jump length displacement histogram (*right*) of **(a)** normal Brownian diffusion and **(b)** Fractional Brownian motion, with α=0.4.

2.2.2 Derivation of the propagator for anomalous diffusion models

Spot-On relies on an analytical description of the propagator in order to perform a reasonably fast fitting. It also relies on tabulated defocalization values that depend on the type of anomalous diffusion model used. The propagators for the main diffusion models are as follows (Weigel et al. 2012,metzler$_{\text{restaurant2004}}$):

- Brownian motion: $P(r,\Delta\tau) = \frac{1}{\sqrt{4\pi D\tau}} e^{-r^2/4Dt}$
- Fractional Brownian motion: $P(r,\Delta\tau) = \frac{1}{\sqrt{4\pi D_\alpha \Delta\tau^\alpha}} e^{-r^2/4D_\alpha \Delta\tau^\alpha}$
- Diffusion inside fractals: $P(r,\Delta\tau) = \frac{d_w \Gamma(d_f/2)}{2\pi^{d_f/2}\Gamma(d_f/d_w)} \left(\frac{1}{4D_F\Delta\tau}\right)^{d_f/d_w} e^{-\frac{r^{d_w}}{4D_F\Delta\tau}}$ (O'Shaughnessy and Procaccia 1985)
- CTRW: $P(r,\Delta\tau) \sim c_1 \Delta\tau^{-\alpha/2} \xi^{-(1-\alpha)/(2-\alpha)} e^{-c_2 \xi^{1/(1-\alpha/2)}}$ with $\xi = |r|/t^{\alpha/2}$

These propagators are 1D propagators. In order to be used in Spot-On, one needs to (1) compute the 2D propagator by performing a radial integration in (x, y) and (2) take into account localization error, by convolving the resulting propagator by a Gaussian of standard deviation σ.

2.2.3 Conclusion

As of today, this new version of Spot-On, with anomalous diffusion, remains a work in progress. Several steps need to be undertaken: (1) the implementation of the models in the least squares fitting routines. (2) The derivation of new sets of defocalization correction coefficients through Monte Carlo simulations and (3) the validation of these models using simulations and experimental data.

Chapter 3

Taking the angles distribution into account

We started in section 1.3 by characterizing diffusion by modeling the radially integrating propagator using a Brownian diffusion model, and a mixture of Brownian motions. We validated this approach using thorough comparisons with a wide range of realistic simulations and with experimental data.

In section IV.2, we proposed a refined characterization by proposing a method to incorporate non-Brownian motion and a mixture of non-Brownian motion in the estimate, making Spot-On (in theory) capable of inferring several types of motion and mixture thereof.

The current approach of Spot-On, however, suffers critical limitations: the estimated propagator is memory-less: it does not take into account the correlation that might exist between translocations. As such, Spot-On should not be seen as a replacement of analysis approaches based on auto-correlation functions (that study the "memory" of the process), but rather as a complement.

We decided to develop an analytical approach in order to incorporate a second dimension in Spot-On: the 2-translocation angular correlation, by computing the angle distribution between consecutive translocations.

3.1 Rationale

The propagator, as computed in Spot-On and other tools, does not use all the information included in SPT data. In particular, angular correlations between jumps are ignored in Spot-On. In biology, few publications focus on angular distributions in a SPT setting (Bouzigues and Dahan 2007; Bhatia et al. 2016). Angular distributions nonetheless carry crucial information about anomalous diffusion, as detailed in (Burov et al. 2013) and used in (Izeddin, Récamier, et al. 2014; Anders S. Hansen, Amitai, et al. 2018). Indeed, at constant localization error, how often a particle "backtracks" is indicative of the type of motion. The computation of the distribution of angles between translocations provides information is then an under-represented metric, and no null model of the the distribution of angles exists under free diffusion.

Under free diffusion, the angle between successive translocations is isotropic. However, when non-zero localization error is taken into account, one expects an angle distribution skewed towards "backward" moves. After a short literature review, we realized that such analytical distribution did not exist.

An estimate of the distribution of angles under a model of Brownian motion with localization error can provide: (1) a more robust estimate of the localization error. (2) a more robust estimate of the diffusion coefficient D and (3) a quantitative indicator of deviations from free diffusion.

Our work has connections with the fields of wrapped distributions and directional statistics (F. Wang and Gelfand 2013; Alan Lee 2010). In particular the distribution of the dot product between two unit vectors has been investigated (Craig 1936; Saw 1983; Nadarajah and Pogány 2016), and specific families of angular distributions, such as wrapped normal distributions and von Mises distributions

(F. Wang and Gelfand 2013; F. Wang and Gelfand 2014; Hernandez-Stumpfhauser, Breidt, and van der Woerd 2017)

3.2 Theory

We first define the process we want to study. It corresponds to a *null model* of SPT data that we currently acquire, namely it models a particle freely diffusing with a diffusion coefficient D, observed under a given localization error σ at a frame rate Δt. This process can be mathematically expressed as a Brownian motion with noise Ξ_t:

$$\Xi_t = \int_{t'=0}^{t} \sqrt{2D}dt' dB_{t'} + \sigma^2 dG_t,$$

with dB_t and dG_t two normal unit Brownian processes. We note that the localization error term (starting with σ^2) does not depend on the framerate: even for an immobile particle ($D = 0$), the particle displays some apparent motion.

Consider three consecutive steps (X, Y, Z) of a 2D Brownian motion sampled at times $\Delta t, 2\Delta t, 3\Delta t)$, that is, for a given time t, $X = \Xi_t$, $Y = \Xi_{t+\Delta_t}$, $Z = \Xi_{t+2\Delta t}$. Following the properties of Brownian motion, it follows:

Figure 3.1: Computing the distribution of angles between two consecutive jumps, U and V, on performed between t and $t + \Delta t$ and the second between $t + \Delta t$ and $t + 2\Delta t$.

$$X \sim \mathcal{N}(0, 2D\Delta t + \sigma^2) \tag{3.1}$$
$$Y \sim \mathcal{N}(0, \sigma^2) \tag{3.2}$$
$$Z \sim \mathcal{N}(0, 2D\Delta t + \sigma^2) \tag{3.3}$$

since the intervals $[0, \Delta t]$ and $[0, 2\Delta t]$ are distinct,

Let $U = Y - X$ and $V = Z - Y$ the two 2D displacement vectors. Consider the couple of 2D random variables $W = (U, V)$:

$$W = \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} U_x \\ U_y \\ V_x \\ V_y \end{pmatrix}, \tag{3.4}$$

and let Σ its covariance matrix:

$$\Sigma = \begin{pmatrix} 1 & 0 & \rho & 0 \\ 0 & 1 & 0 & \rho \\ \rho & 0 & 1 & 0 \\ 0 & \rho & 0 & 1 \end{pmatrix} \left(2D\Delta t + 2\sigma^2\right) \text{ , with } \rho = \frac{-4\sigma^2}{2D\Delta t + 2\sigma^2} \tag{3.5}$$

Since the vector (U_x, U_y, V_x, V_y) is a Gaussian vector, its multi-dimensional density is given in Cartesian coordinates by:

$$f_W^{cart}\left(u_x, u_y, u_x, u_y\right) = \frac{1}{\sqrt{(2\pi)^4|\Sigma|}} e^{-\frac{1}{2}W^T\Sigma^{-1}W}$$

From this general expression, we want to express the dependency of the density on the relative angle between the vectors U and V. Let r_1 and θ_1 (resp. r_2 and θ_2) the parameters of the polar decomposition of U (resp. V) in the 2D plane:

$$w = \begin{pmatrix} u_x \\ u_y \\ v_x \\ v_y \end{pmatrix} = \begin{pmatrix} r_1\cos\theta_1 \\ r_1\sin\theta_1 \\ r_2\cos\theta_2 \\ r_2\sin\theta_2 \end{pmatrix} \tag{3.6}$$

and compute $f_W^{cart}(w) = f_W^{cart}\left(r_1\cos(\theta_1), r_1\sin(\theta_1), r_2\cos(\theta_2), r_2\sin(\theta_2)\right)$. One first derives:

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{1-\rho^2} & 0 & \frac{\rho}{\rho^2-1} & 0 \\ 0 & \frac{1}{1-\rho^2} & 0 & \frac{\rho}{\rho^2-1} \\ \frac{\rho}{\rho^2-1} & 0 & \frac{1}{1-\rho^2} & 0 \\ 0 & \frac{\rho}{\rho^2-1} & 0 & \frac{1}{1-\rho^2} \end{pmatrix}, \tag{3.7}$$

and $|\Sigma| = 1 - 2\rho^2 + \rho^4$. One can then express: $w^T\Sigma^{-1}w$:

$$w^T\Sigma^{-1}w = \frac{1}{1-\rho^2}\left(r_1^2 + r_2^2 - 2\rho r_1 r_2\cos(\theta_1 - \theta_2)\right)$$

One notes that the expression only depends on the relative angle between U and V $(\theta_1 - \theta_2)$. This expression naturally leads to the distribution of angles, that we express as a function of $(r_1, r_2, \theta_1 - \theta_2)$:

$$f_W^{polar}\left(r_1, r_2, \theta_1 - \theta_2\right) = \frac{1}{\sqrt{(2\pi)^4(1-2\rho^2+\rho^4)}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(r_1^2 + r_2^2 - 2\rho r_1 r_2\cos(\theta_1 - \theta_2)\right)\right]$$

3.3 Detailed calculations

3.3.1 Computation of the density

Let (x, y, z, t) as a simplified notation: $w^T = \left(r_1\cos\theta_1, r_1\sin\theta_1, r_2\cos_2^\theta, r_2\sin\theta_2\right)^T = (x, y, z, t)^T$ and compute $w^T\Sigma^{-1}w$. We remind that:

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & 0 & -\rho & 0 \\ 0 & 1 & 0 & -\rho \\ -\rho & 0 & 1 & 0 \\ 0 & -\rho & 0 & 1 \end{pmatrix} \tag{3.8}$$

It then follows:

$$w^T\Sigma^{-1}w = \frac{1}{1-\rho^2}(x,y,z,t)\,(x-\rho z, y-\rho t, z-\rho x, t-\rho y)^T \tag{3.9}$$

$$= \frac{1}{1-\rho^2}\left(x(x-\rho z)+y(y-\rho t)+z(z-\rho x)+t(t-\rho y)\right) \tag{3.10}$$

$$= \frac{1}{1-\rho^2}\left(x^2+y^2+z^2+t^2-\rho(xz+yt+zx+ty)\right) \tag{3.11}$$

Reverting to original notations, one then notices that:

$$\left\{\begin{array}{rcl} x^2+y^2 & = & (r_1\cos\theta_1)^2+(r_1\sin\theta_1)^2 = r_1^2 \\ z^2+t^2 & = & (r_2\cos\theta_2)^2+(r_2\sin\theta_2)^2 = r_2^2 \end{array}\right. \tag{3.12}$$

leading to:

$$w^T\Sigma w = \frac{1}{1-\rho^2}\left(r_1^2+r_2^2-2\rho(xz+yt)\right) \tag{3.13}$$

$$= \frac{1}{1-\rho^2}\left(r_1^2+r_2^2-2\rho(r_1r_2\cos\theta_1\cos\theta 2+r_1r_2\sin\theta_1\sin\theta_2)\right) \tag{3.14}$$

$$= \frac{1}{1-\rho^2}\left(r_1^2+r_2^2-2\rho r_1r_2\cos(\theta_1-\theta_2)\right). \tag{3.15}$$

3.3.2 Case $\rho = 0$

The case $\rho = 0$ arises when there is no localization error $\sigma = 0$. In that case, the angles distribution density reduces to:

$$f_W^{polar}(r_1, r_2, \theta_1-\theta_2|\rho=0) = \frac{1}{2\pi}\exp\left[-\frac{r_1^2+r_2^2}{2}\right]$$

3.4 Conclusion

This result provides an analytical expression for the distribution of angles between two consecutive jumps undergoing Brownian motion, under a non-zero localization error.

Further steps include to perform simulations to determine how quickly an estimator based on the distribution of angles can be estimated, and to develop fitting routines.

Chapter 4

Compressed sensing

4.1 Introduction

In parallel of our work on SPT, we developed another imaging approach, based on the technique of compressed sensing. Indeed, many objects need to be imaged in 3D at a fast rate, and often the amount of light needed to perform such acquisitions causes significant photodamage/phototoxicity.

In order to circumvent this issue, we proposed an imaging scheme that assumes some characteristics of the image (the type of observed sample is known) in order to acquire less information. The images are acquired at high speed using a so-called "compressed" imaging scheme and decompressed using sparse reconstruction algorithms, with the knowledge of the type of object to be observed.

4.2 Publication

Research Article | Vol. 25, No. 12 | 12 Jun 2017 | OPTICS EXPRESS 13668
Optics EXPRESS

Faster and less phototoxic 3D fluorescence microscopy using a versatile compressed sensing scheme

MAXIME WORINGER,[1,2,3,4,5] XAVIER DARZACQ,[2] CHRISTOPHE ZIMMER[1,4,5,6] AND MUSTAFA MIR[2]

[1] *Unité Imagerie et Modélisation, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France*
[2] *Department of Molecular and Cell Biology, University of California, Berkeley, USA*
[3] *Sorbonne Universités, UPMC Univ Paris 06, IFD, 4 Place Jussieu, 75252 Paris cedex 05, France*
[4] *Centre National de la Recherche Scientifique (CNRS), UMR 3691, Paris, France*
[5] *Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), USR 3756, Institut Pasteur et CNRS, Paris, France*
[6] *czimmer@pasteur.fr*
[*] *mmir@berkeley.edu*

Abstract: Three-dimensional fluorescence microscopy based on Nyquist sampling of focal planes faces harsh trade-offs between acquisition time, light exposure, and signal-to-noise. We propose a 3D compressed sensing approach that uses temporal modulation of the excitation intensity during axial stage sweeping and can be adapted to fluorescence microscopes without hardware modification. We describe implementations on a lattice light sheet microscope and an epifluorescence microscope, and show that images of beads and biological samples can be reconstructed with a 5-10 fold reduction of light exposure and acquisition time. Our scheme opens a new door towards faster and less damaging 3D fluorescence microscopy.

OCIS codes: (110.1758) Computational imaging; (170.2520) Fluorescence microscopy; (170.6900) Three-dimensional microscopy.

References and links

1. P. Keller, A. Schmidt, J. Wittbrodt, and E. Stelzer,"Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy," Science **322**, 5904 (2008).
2. T. A. Planchon, L. Gao, D. E. Milkie, M. W. Davidson, J. A. Galbraith, C. G. Galbraith, and E. Betzig, "Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination," Nat. Methods **8**, 417–423 (2011).
3. M. Weber, M. Mickoleit, and J. Huisken, "Light sheet microscopy," in "*Methods in Cell Biology*" (Elsevier, 2014), pp. 193–215.
4. R. Baraniuk, "Compressive sensing," IEEE Signal Processing Mag. **24**, 118–121 (2007)
5. D. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**, 1289–1306 (2006).
6. E. Candes and M. Wakin, "An introduction to compressive sampling," IEEE Signal Processing Mag. **25**, 21–30 (2008).
7. Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications* (Cambridge University, 2012).
8. M. Elad, *Sparse and Redundant Representations* (Springer, 2010).
9. J. Bobin, J.-L. Starck, and R. Ottensamer, "Compressed sensing in astronomy," IEEE J. Sel. Top. Signal Process **2**, 718–726 (2008).
10. M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing MRI," IEEE Signal Processing Mag. **25**, 72–82 (2008).
11. J. Hunt, T. Driscoll, A. Mrozack, G. Lipworth, M. Reynolds, D. Brady, and D. R. Smith, "Metamaterial apertures for computational imaging," Science **310**, 399 (2013).
12. J. Shin, B. T. Bosworth, and M. A. Foster, "Compressive fluorescence imaging using a multi-core fiber and spatially dependent scattering," Opt Lett **109**, 42 (2017).
13. L. Gao, J. Liang, C. Li, and L. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," Nature **516**, 74–77 (2014).
14. J. Liang, C. Ma, L. Zhu, Y. Chen, L. Gao, and L. V. Wang, "Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse," Sci Adv **3**, e1601814 (2017).
15. E. McLeod and A. Ozcan, "Unconventional methods of imaging: computational microscopy and compact implementations," Rep. Prog. Phys. **79**, 076001 (2016).

#292516 https://doi.org/10.1364/OE.25.013668
Journal © 2017 Received 13 Apr 2017; revised 16 May 2017; accepted 19 May 2017; published 7 Jun 2017

Optics EXPRESS

16. M. M. Marim, E. D. Angelini, and J.-C. Olivo-Marin, "A compressed sensing approach for biological microscopic image processing," in "2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro," (IEEE, 2009), pp. 1374–1377.
17. P. Ye, J. L. Paredes, Y. Wu, C. Chen, G. R. Arce, and D. W. Prather, "Compressive confocal microscopy: 3d reconstruction algorithms," in "SPIE MOEMS-MEMS: Micro-and Nanofabrication," (International Society for Optics and Photonics, 2009), pp. 72100G.
18. Y. Wu, P. Ye, I. O. Mirza, G. R. Arce, and D. W. Prather, "Experimental demonstration of an optical-sectioning compressive sensing microscope (CSM)," Opt. Express **18**, 24565–24578 (2010).
19. V. Studer, J. Bobin, M. Chahid, H. S. Mousavi, E. Candes, and M. Dahan, "Compressive fluorescence microscopy for biological and hyperspectral imaging," Proc. Natl. Acad. Sci. U. S. A. **109**, E1679–E1687 (2012).
20. S. Schwartz, A. Wong, and D. A. Clausi, "Compressive fluorescence microscopy using saliency-guided sparse reconstruction ensemble fusion," Opt. Express **20**, 17281–17296 (2012).
21. L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," Nat. Methods **9**, 721–723 (2012).
22. N. Pavillon and N. I. Smith, "Compressed sensing laser scanning microscopy," Opt. Express **24**, 30038 (2016).
23. E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," Appl. Comput. Harmon Anal. **31**, 59–73 (2011).
24. T. Tao and E. Candès, "Decoding by linear programming," arXiv preprint arXiv:math/0502327 (2004).
25. M. Raginsky, R. M. Willett, Z. T. Harmany, and R. F. Marcia, "Compressed sensing performance bounds under Poisson noise," IEEE Trans. Signal Process **58**, 3990–4002 (2010).
26. E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**, 489–509 (2006).
27. Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms–theory and practice," IEEE Trans. Signal Process **21**, 1084–1096 (2012).
28. S. Becker, J. Bobin, and E. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," arXiv preprint arXiv:0904.3367 (2009).
29. R. Baraniuk, M. A. Davenport, M. F. Duarte, and C. Hegde, *An Introduction to Compressive Sensing* (OpenStax CNX, 2011).
30. C. Bilen, Y. Wang, and I. Selesnick, "Compressed sensing for moving imagery in medical imaging," arXiv preprint arXiv:1203.5772 (2012).
31. B.-C. Chen, W. R. Legant, K. Wang, L. Shao, D. E. Milkie, M. W. Davidson, C. Janetopoulos, X. S. Wu, J. A. Hammer, Z. Liu, B. P. English, Y. Mimori-Kiyosue, D. P. Romero, A. T. Ritter, J. Lippincott-Schwartz, L. Fritz-Laylin, R. D. Mullins, D. M. Mitchell, J. N. Bembenek, A.-C. Reymann, R. Bohme, S. W. Grill, J. T. Wang, G. Seydoux, U. S. Tulu, D. P. Kiehart, and E. Betzig, "Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution," Science **346**, 1257998 (2014).
32. A. D. Edelstein, M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, and N. Stuurman, "Advanced methods of microscope control using micro-Manager software," J. Biol. Methods **1**, 10 (2014).
33. J.-Y. Tinevez, J. Dragavon, L. Baba-Aissa, P. Roux, E. Perret, A. Canivet, V. Galy, and S. Shorte, "A quantitative method for measuring phototoxicity of a live cell imaging microscope," in "*Methods in Enzymology*" (Elsevier, 2012), pp. 291–309.
34. D. S. Smith, J. C. Gore, T. E. Yankeelov, and E. B. Welch, "Real-time compressive sensing MRI reconstruction using GPU computing and split Bregman methods," Int. J. Biomed. Imaging **2012**, 1–6 (2012).
35. B. E. Nett, J. Tang, and G.-H. Chen, "GPU implementation of prior image constrained compressed sensing (PICCS)," Proc. SPIE **7622**, 762239 (2010).
36. M. Mordechay and Y. Y. Schechner, "Matrix optimization for Poisson compressed sensing," in "2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)" (IEEE, 2014), pp. 684–688.
37. J. D. Hunter, "Matplotlib: A 2d graphics environment," Comput. Sci. Eng. **9**, 90–95 (2007).

1. Introduction

Imaging fluorescently labeled biological structures with high spatio-temporal resolution requires judicious compromises between the conflicting goals of achieving high signal-to-noise ratio (SNR) and temporal resolution while keeping the excitation power low to minimize photobleaching and phototoxicity. For example, to obtain a higher SNR, one can either increase the exposure time, thereby reducing imaging speed, or increase the illumination power, thereby increasing photodamage. These tradeoffs are further exacerbated in 3D imaging, which is often required in biological applications, such as calcium imaging in neurons or transient mitotic events in a developing embryo. Traditionally, 3D microscopy images are obtained by sequentially acquiring 2D images of individual focal planes, where axial spacing is dictated by the Nyquist sampling criterion to achieve optimal spatial resolution in all dimensions. As a consequence, hundreds or

Optics EXPRESS

thousands of planes are needed to image samples 10-1,000 μm thick, dramatically increasing acquisition time and light exposure. Although light sheet illumination considerably reduces photodamage and allows prolonged imaging of living cells in 3D, the required hardware systems are often costly and scarce, and the acquisition time for each volume still remains constrained by the Nyquist criterion [1–3].

The field of compressed sensing, introduced over a decade ago, offers an avenue to overcome these limitations [4–7]. Compressed sensing leverages the fact that natural images are highly non-random and harbour intrinsic redundancies, which can be formulated as sparsity in an appropriate linear basis [8]. This sparsity can be exploited in order to reconstruct images from fewer measurements than specified by the Nyquist criterion, provided that measurements are taken in an appropriate manner. Compressed sensing has been successfully applied in diverse imaging applications in fields including astronomy [9], magnetic resonance imaging [10], lensless imaging [11, 12] and ultrafast imaging [13, 14], where it has enabled a considerable increase in acquisition speeds.

In biological microscopy, compressed sensing should in principle enable similar benefits in reducing acquisition time and light exposure without compromising SNR [15]. However, despite several proof of concepts, fluorescence microscopy has benefited relatively little from compressed sensing approaches in practice. One reason for this is that most compressed sensing strategies proposed to date require considerable modifications of the optical system, an important impediment for application on routinely used microscopes [16–21]. We note however, that a compressed sensing scheme without modification of the light path was recently used in confocal laser scanning microscopy to achieve a 10-15 fold speedup in 2D imaging [22].

Here we introduce a compressed sensing scheme for 3D fluorescence imaging that relies on compression along the optical axis (z axis) and is applicable to a large range of fluorescence modalities without modification of the optical path. We show that for a given SNR, our method can reconstruct a z stack from a 2-10 times faster acquisition than traditional plane-by-plane imaging with Nyquist sampling. For dynamic microscopy of live samples, this approach opens the door to either lower excitation power and photodamage (at constant acquisition speed and SNR) or to higher temporal resolution (at constant excitation power and SNR).

In Section 2, we first describe our method conceptually, starting with a brief reminder of the basics of compressed sensing. In Section 3, we present results on simulations. Implementation on a lattice light sheet microscope and a conventional epifluorescence microscope are demonstrated in Sections 4 and 5 respectively. Sections 6 and 7 provide a brief discussion and conclusion.

2. Method

2.1. Basics of compressed sensing

Compressed sensing is based on the realization that under certain (broad) conditions, natural signals such as images can be reconstructed from a smaller number of measurements than prescribed by Nyquist sampling. If X is a (vectorized) image of size $N \times 1$ and $\mathbf{A}$ a known $M \times N$ matrix that transforms X into a signal $Y = \mathbf{A}X$ of smaller size $M \times 1$ ($M < N$), then the goal is to recover X (or a good approximation thereof) from Y. The matrix $\mathbf{A}$, which is independent of the data, specifies how the N pixels of the image are scrambled into the M "compressed" measurements and is called the sensing (or measurement) matrix.

In order to recover X from Y, compressed sensing reconstruction algorithms exploit the structural redundancy of images. In the simplest setting, it is assumed that the image X is sparse, i.e. that the number of non-zero values, $K = ||X||_{\ell_0}$ is small ($K << N$), or that X can be represented sparsely in a suitable basis, i.e. that $X = \Psi\alpha$, where Ψ is an invertible (e.g. orthonormal) $N \times N$ matrix and α is a sparse vector of size $N \times 1$. Note that this setting can easily be adapted to incorporate a redundant dictionary Ψ of size $W \times N$ with $W > N$ instead of an orthogonal basis, allowing for improved reconstructions [23]. The reconstruction algorithms

Optics EXPRESS

aim to determine the sparsest representation α consistent with the data, i.e. such that $X = \mathbf{A}\Psi\alpha$. While minimizing the ℓ_0 norm to enforce sparsity is NP-hard and computationally unfeasible, minimizing the ℓ_1 norm ($||\alpha||_{\ell_1} = \Sigma_{i=1}^{N}|\alpha_i|$) leads to computationally tractable optimization algorithms that recover the exact solution.

In practice, images are noisy and only approximately sparse, therefore compressed sensing algorithms seek to recover approximations of X by determining the sparsest representation α such that $Y \approx \mathbf{A}\Psi\alpha$. For additive gaussian noise, this is typically done by solving the optimization problem : $\alpha^* = \arg\min_\alpha F(\alpha)$ for objective functions $F(\alpha)$ such as:

$$F(\alpha) = ||\alpha||_{\ell_1} + \lambda||\mathbf{A}\bar{}\alpha - Y||_{\ell_2}^2 \tag{1}$$

where ℓ_2 is the Euclidian norm and λ a Lagrange multiplier.

Under suitable conditions for $\mathbf{A}$, such as the restricted isometry property (which is fulfilled in particular for random Gaussian matrices), it was shown that a good approximation of the N values in α can be recovered from a number of compressed measurements $M = O\left(K\log(N/K)\right)$, which can be much smaller than N [6,24–26]. The reconstructed image is then simply obtained as $X^* = \Psi\alpha^*$. For images corrupted by Poisson noise, the appropriate objective function becomes:

$$F(\alpha) = ||\alpha||_{\ell_1} + \lambda\mathcal{L}(\alpha) \tag{2}$$

and its minimization is subject to the positivity constraint: $\Psi\alpha \geq 0$, where $\mathcal{L}(\alpha)$ designates the negative Poisson log-likelihood $\mathcal{L}(\alpha) = \mathbb{1}^T\mathbf{A}\Psi\alpha - \Sigma_{i=1}^{M}Y_i\log\left(e_i^T\mathbf{A}\Psi\alpha\right)$, $\mathbb{1}$ is a $M \times 1$ vector of ones, e_i is the canonical basis vector i and the T superscript denotes transposition [27]. A variety of efficient algorithms for compressed sensing recovery have been proposed, mostly for Gaussian noise, but also for Poisson noise [27, 28]. See [7, 8, 29] for in-depth introductions to sparsity and compressed sensing.

2.2. Axially compressed imaging scheme

The traditional way to image a 3D volume is to successively scan the focal plane of the microscope along the z axis in a step-wise fashion, with spacing Δz, and acquire a 2D image at each $z = k\Delta z$ position ($k = 1 \ldots N$). We hereafter refer to this imaging scheme as *plane-by-plane* acquisition, see Fig. 1(a). In this scheme, the focal plane position is given by:

$$z_f(t) = E\left(\frac{t}{\Delta t}\right) \times \Delta z$$

where Δt is the camera exposure time and $E(x)$ is the next smallest integer to x. The spacing Δz is usually dictated by the point spread function (PSF) width along the z-axis (Nyquist sampling). In plane-by-plane imaging, the k-th camera frame, $F_k^{p.-by-p.}$, $k = 1 \ldots N$ carries information from the $z = k\Delta z$-plane only and is given by:

$$F_k^{p.-by-p.}(x, y) = L_0 \times \Delta t \times (I * PSF)(x, y, k\Delta z) \qquad \textbf{for } k = 1 \ldots N \tag{3}$$

where $I(x, y, z)$ designates the 3D distribution of fluorophores in the sample, L_0 is the laser intensity, PSF is the 3D PSF of the microscope and $*$ stands for convolution. In this setting, the acquisition time for a full 3D z-stack with N focal planes is $N\Delta t$ and the light dose received by the sample is $NL_0\Delta t$.

In the compressed sensing imaging scheme proposed here, the axial dimension of a 3D stack is acquired in a compressed fashion, such that the k-th acquired frame no longer contains information from the $z = k\Delta z$ position only, but is a linear combination of information from multiple z-planes:

Optics EXPRESS

$$F_k^{comp}(x,y) = L_0 \times \Delta t \times \int_0^{(N-1)\Delta z} A_k(z) \times (I * PSF)(x,y,z)dz \qquad \textbf{for } k = 1 \ldots M < N \quad (4)$$

where $A_k(z)$ is a function that describes how the image intensity profile along the z-axis is combined into the single frame k. Note that this scheme is compressed in the sense that it requires $M < N$ frames. This expression can be approximated and discretized as:

$$F_k^{comp}(x,y) = L_0 \times \Delta t \times \sum_{i=1}^{i=N} \mathbf{A}_{k,i} \times (I * PSF)(x,y,i\Delta z) \qquad \textbf{for } k = 1 \ldots M < N \quad (5)$$

In Eq. 5 above, the matrix $\mathbf{A}$ is the discrete counterpart of $A_k(z)$ $(k = 1 \ldots M)$ and describes how the image intensity from each of the N z-planes of the stack is combined into a single value. Note that in the case where the measurement matrix is the identity matrix ($A_{k,i} = \delta(i,k)$ with $N = M$ and δ the Kronecker symbol), this scheme reduces to the plane-by-plane imaging scheme of Eq. 3, see Fig. 1(b).

The connection with the compressed sensing setting outlined in Section 2.1 is immediate. For any fixed (x,y) location, $Y_k = F_k^{comp}(x,y)$ $(k = 1 \ldots M)$ defines a $M \times 1$ vector Y, that is a set of compressed measurements obeying $Y = \mathbf{A}X$, where $X = \tilde{I}$ is the $N \times 1$ vector corresponding to the intensity profile along the z axis convolved with the PSF and sampled every Δz, i.e.: $\tilde{I}_i = (I * PSF)(x,y,i\Delta z)$ for $i = 1 \ldots N$. Thus, the matrix formulation of compressed sensing for a given (x,y) location is: $F^{comp} = \mathbf{A}\tilde{I}$. Applying the results mentioned in Section 2.1, under suitable conditions, approximate recovery of $\tilde{I}$ should therefore be possible from the compressed measurements F^{comp}.

2.3. Physical implementation

In practice, our microscopy system achieves axial compression in the following way. During each camera exposure, the stage is swept at constant speed across the entire z-range of the volume to be imaged, see Fig. 1(c):

$$z_f(t) = N\Delta z \left(\frac{t}{\Delta t} - E\left(\frac{t}{\Delta t}\right) \right)$$

In this way, each pixel (x,y) of the camera records an integration of the emitted fluorescence along the z axis. During this axial sweep, the excitation laser intensity at the sample, $L(t) = L_0 T(t)$ is modulated over time, such that:

$$\begin{aligned} F_k^{comp}(x,y) &= L_0 \int_{(k-1)\Delta t}^{k\Delta t} T(t) \times (I * PSF)(x,y,z_f(t))\, dt \\ &= \frac{\Delta t}{N\Delta z} L_0 \int_0^{(N-1)\Delta z} T(z) \times (I * PSF)(x,y,z)\, dz \end{aligned}$$

is an average of the fluorescence distribution along z weighted by the modulated excitation light intensity. In the discrete approximation, we have:

$$F_k^{comp}(x,y) = L_0 \times \Delta t \times \frac{1}{N} \sum_{i=1}^{i=N} T_{k,i} \times (I * PSF)(x,y,i\Delta z) \quad (6)$$

where $L_0 \times T_{k,i}$ is the laser power applied during frame k when the focal plane is at $z = i\Delta z$. This modulation obeys a user-defined pattern specified by the k-th row of the measurement matrix $\mathbf{A}$, as in Eq. 5, see Fig. 1(c), i.e. we set $\mathbf{T} = N\mathbf{A}$.

Optics EXPRESS

This procedure is repeated for all M rows of the measurement matrix, resulting in M compressed 2D images $F_1^{comp}, \ldots F_M^{comp}$. For a given constant exposure time Δt, the acquisition speedup compared to plane-by-plane imaging is thus simply given by the ratio: $\kappa = N/M$. Figure 1(c) jointly shows the z position of the stage and the illumination intensity for a measurement matrix $\mathbf{A}$ consisting of a truncated Fourier basis shown in Fig. 1(d).

A major advantage of this scheme is that it can be implemented without modification of the light path and simply requires a synchronization of two microscope components: (1) the z-piezo, which allows precise control of the axial focus (sample or objective) and (2) the AOTF (acousto-optic tunable filter) which allows to precisely and rapidly modulate the excitation light intensity transmitted to the sample ($\mathbf{T}$). Instead of using an AOTF, it is in principle possible to modulate the intensity of the light source directly.

Note, that since the optical coding is performed by light modulation, this setup leaves absolute freedom of choice for the measurement matrix $\mathbf{A}$, as long as its values are all positive. For example, random sensing matrices, which allow compressed sensing reconstruction for images sparse in any transform basis Ψ (universality property), can be implemented in a straightforward manner. In this paper, we choose a Fourier matrix, which is an optimal sensing matrix for images that are sparse in the direct spatial domain, taking the M first rows of the matrix, from low to high frequencies, see Fig. 1(d). Importantly, we linearly scale the matrix $\mathbf{A}$ such that all its values fall between 0 and $1/N$. This ensures that during each frame k, the sample receives a light dose of $L_0 \Delta t N^{-1} \sum_{i=1}^{N} T_{k,i} = L_0 \Delta t \sum_{i=1}^{N} A_{k,i} \leq L_0 \Delta t N \frac{1}{N} = L_0 \Delta t$, i.e. less or equal to the dose received in plane-by-plane imaging. Therefore, the total light dose received by the sample during a compressed imaging acquisition with M frames is at least $\kappa = N/M$ times less than in the plane-by-plane acquisition, where κ is the compression ratio with respect to the plane-by-plane, Nyquist sampling.

2.4. Sparsity prior and PSF model

In this paper, we assume for simplicity that the 3D distribution of fluorescent structures is sparse in the spatial domain, but we take into account the 3D blurring caused by diffraction. This is done by incorporating a model of the 3D PSF into the $W \times N$ redundant dictionary Ψ, such that the 3D image can be modeled as: $X = \Psi\alpha$, where α represents the 3D distribution of fluorescent structures and is assumed to be (approximately) sparse. In practice, we first measure the empirical PSF of the microscope using a conventional plane-by-plane z-stack and derive one cropped 2D image in the (x, z) plane. We then build the dictionary Ψ: each element of the dictionary is defined as a translation in the (x, z) plane of the empirical PSF within a given 2D reconstruction window. Thus, the dictionary is a collection of PSFs at various locations. Before reconstruction, both the compressed stacked and the elements of the dictionary are flattened into 1D vectors. Therefore, although our scheme performs compression only along the z axis, the reconstruction algorithm incorporates a 2D sparsity prior [30].

2.5. Numerical implementation

The reconstructions are performed using a custom port in Python of the previously published `SPIRAL-TAP` algorithm [27], which solves the optimization problems (1) or (2) above. Our port is publicly available, together with sample data and analysis scripts (see Section Software and data availability).

In practice, performing a reconstruction on a full 3D (or even 2D) image is not computationally tractable. Since we use a PSF model with a restricted spatial support, we assume that two (x, y) positions located farther apart than the characteristic width of the PSF are independent and reconstruct them in parallel on a computing cluster. We then calculate a single 2D image by averaging small overlapping chunks, and stack them together to obtain a 3D image.

Optics EXPRESS

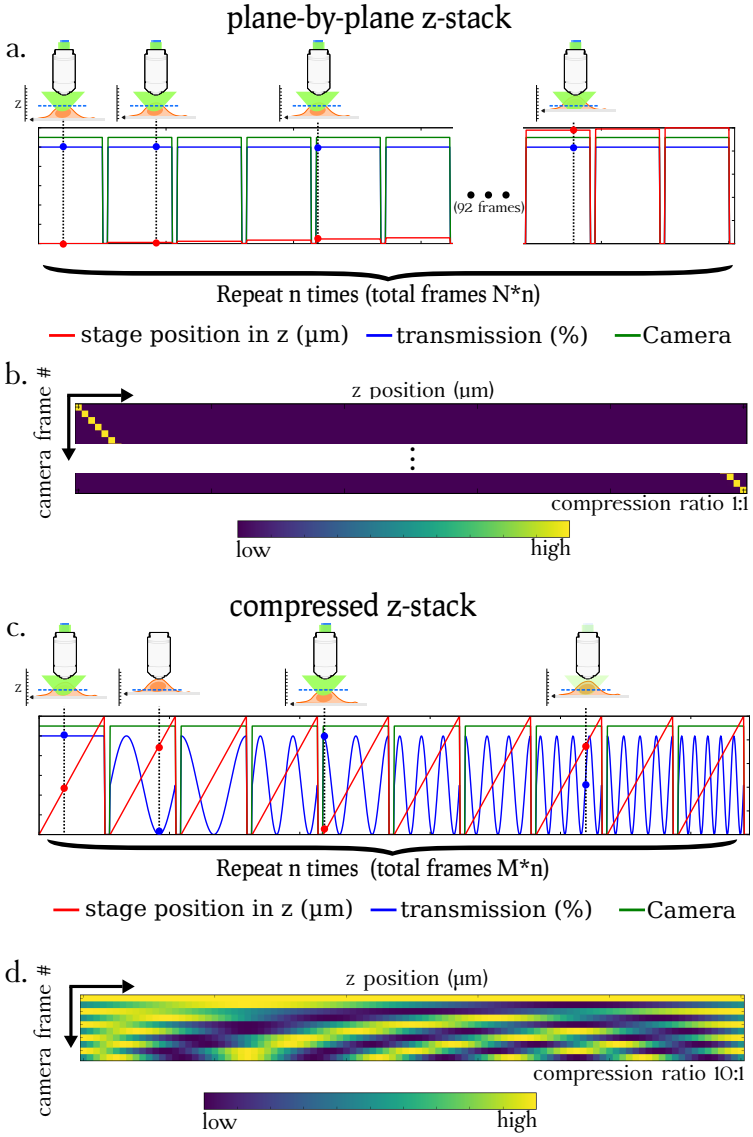


Fig. 1. Principle of the proposed 3D compressed imaging method compared to traditional 3D plane-by-plane imaging. **(a)** Plane-by-plane imaging: for each camera frame (green curve indicates if the shutter is open or closed), one plane of the sample (red curve indicates z position) is illuminated at a constant laser intensity (blue curve shows transmission percentage of the excitation light source). The process is repeated for each plane ($N = 101$ times) to acquire a z-stack. Finally, the full imaging sequence is repeated n times to acquire a 4D movie for a total of $N * n$ frames. The blue and red dots represent the illumination intensity and stage position at each time point respectively. This imaging scheme can be represented as the application of a square diagonal measurement matrix **A** as shown in **(b)**: for each camera frame (row), only one z plane is illuminated (column). **(c)** Axially compressed imaging: the stage continually sweeps through the entire axial range while the illumination is modulated to create a specific axial light pattern. In this scheme, multiple planes of the sample are illuminated during a single camera exposure frame. This process is repeated $M = 10 < N$ times with different light patterns, thus performing an opto-mechanical implementation of a compressed measurement matrix, as shown in **(d)**. Finally, the full imaging sequence is repeated n times to acquire a 4D dataset with a total of M*n (10*101) frames.

Optics EXPRESS

3. Compressed imaging on simulated images

3.1. Generation of test images and metrics

We first tested our compressed sensing approach on simulated images. For this purpose, we generated a series of 100 synthetic images, termed ground truth (**GT**). The images contained features at different scales and sparsity levels (see Fig. 2(a, top)) along the compression axis (z axis). The images are scaled so that the pixel of highest intensity had a value of $I_{max} = 10,000$ counts. For simplicity, the effects of diffraction blurring is ignored here.

Next, to simulate an acquisition in realistic conditions in the conventional plane-by-plane scheme, we corrupted the ground truth images **GT** using Poisson noise ($\mathcal{P}$) and additive half-normal noise ($|\mathcal{N}|$). This resulted in noisy reference images ($\mathbf{NR_{SNR}}$) for different SNRs (see below): $\mathbf{NR_{SNR}} = \mathcal{P}(\mathbf{GT}) + |\mathcal{N}|(\sigma\sqrt{\pi/2})$ where σ corresponds to the expected number of background photons to achieve a given SNR, that is: $\sigma = I_{max}/SNR$. We also simulated images acquired by reducing the exposure time per frame by a factor of ten, allowing us to compare the plane-by-plane and the compressed acquisition scheme at a constant acquisition time per z-stack. To do so, we assumed that the SNR scales with the exposure time Δt as $SNR \propto \sqrt{\Delta t}$ and simulated plane-by-plane images with the corresponding SNR. We denote those images as $\mathbf{NR^{10x}_{SNR}}$.

In parallel, we computed compressed versions of the same ground truth image by applying the measurement matrix **A** to **GT** (for different compression ratios κ, i.e. varying numbers of rows of **A**), and subsequently applied Poisson and additive Gaussian noise as for the images in the plane-by-plane acquisition ($F^{comp} = \mathcal{P}\left(\mathbf{A}\tilde{I}\right) + |\mathcal{N}|(\sigma\sqrt{\pi/2})$).

Then, we computed reconstructions from these noisy compressed images, and denote the resulting 3D images as $\mathbf{CS^{\kappa}_{SNR}}$. The reconstructions were performed without a PSF model, thus assuming sparsity of the reconstructed image and using $\Psi = \mathbf{I_N}$. The SNR was computed as the mean of the non-zero pixels of the ground truth image divided by the mean of the additive noise, see Fig. 2(b).

Finally, we quantified reconstruction quality by computing the mean square error $MSE(\mathbf{GT}, \mathbf{CS^{\kappa}_{SNR}})$ between the ground truth images **GT** and the compressed sensing reconstructions $\mathbf{CS^{\kappa}_{SNR}}$. For comparison, we also computed the MSE between the ground truth images and the plane-by-plane acquisition image obtained for the same SNR, $MSE(\mathbf{GT}, \mathbf{NR_{SNR}})$ and with a ten times lower exposure time ($MSE(\mathbf{GT}, \mathbf{NR^{10x}_{SNR}})$).

3.2. Results on simulated images

We first compare reconstructions for a fixed SNR=20 and increasing compression ratios κ, see Fig. 2(b). At low compression ratios (1:2, corresponding to a two fold speedup compared to the plane-by-plane acquisition), all features of the simulated images are accurately reconstructed, including both high and low frequency details. Minor artifacts are visible in the regions of moderate sparsity (arrow x_1). As the compression ratio increases, fine features are progressively lost (arrow x_2), whereas larger objects remain visible at their correct location (arrows x_1 and x_3). At high compression ratios, the reconstructed intensity significantly diverges from the ground truth. Nevertheless, this first example illustrates that object positions and shapes can be approximately reconstructed from compressed images with high compression ratios, even when the ground truth images exhibit quite variable levels of sparsity.

We then evaluate the influence of different noise levels on reconstruction quality by computing the MSE for SNR ranging from 1 to 80 and for compression ratios ranging from 1:2 to 1:30, see Fig. 2(c). At high SNR, the noisy plane-by-plane reference ($\mathbf{NR_{SNR}}$) exhibits a much lower MSE than the compressed sensing reconstruction. However, as mentioned in the introduction, compressed sensing is most useful for conditions in which photodamage and/or acquisition speed are limiting, i.e. for low SNR images. Figure 2(c) shows that for low SNR (≤ 15), the

Optics EXPRESS

MSE of the noisy reference and the reconstruction become close. Furthermore, in this SNR range, the reconstruction from a 1:10 compression ratio shows a significantly lower MSE than a plane-by-plane acquisition performed in the same overall time, see the dash-dotted line in 2(c). This suggests that our compressed sensing approach can recover images of similar quality as plane-by-plane imaging, but with a considerable reduction of acquisition time and light exposure, and with a higher quality than simply decreasing the exposure time in the traditional plane-by-plane-mode. Equivalently, using the same acquisition time and light exposure as in plane-by-plane imaging, compressed imaging enables reconstructions of higher quality images as measured by the MSE metric.

We note that the reconstructions from a higher compression ratio tend to have a lower MSE than the ones from a higher compression ratio, a result that does not match the visual quality of images shown in Fig. 2(b).This is due to the smoothing (loss of high frequency content) that is observed in the higher compression reconstructions.

Thus, our simulation results suggest that axially compressed sensing acquisition might be a worthwhile alternative to plane-by-plane imaging for faster and less phototoxic 3D microscopy.

4. Compressed lattice light sheet imaging

4.1. Lattice light sheet implementation

Having demonstrated our technique on simulated images, we implemented our compressed sensing scheme on a lattice light sheet microscope (LLSM). In a light sheet microscope [31], one objective is used to produce a very thin sheet of light that illuminates the sample at a 90° angle with the axis of another objective used for detection. Due to its excellent axial resolution, a lattice light sheet microscope is an ideal candidate to implement our compressed sensing scheme.

In LLSM, the focus is adjusted by moving the light sheet (using a scanning galvanometer mirror) and the focus of the 20x/1.1 NA water immersion observation objective Nikon MRD77220 (which is mounted on a piezo stage) in a synchronized manner across the sample, see Fig. 3(a). For the compressed sensing acquisitions, we modified the software generating the FPGA control command (Coleman Technologies) in order to synchronize the motion of the sheet and the observation objective with a custom light modulation produced by the AOTF (AOTFnC-400-650-TN, AA Optoelectronics) during a single camera exposure. Our modified software also allows to load a predefined measurement matrix $\mathbf{A}$ and to set exposure parameters.

For the experiments reported below, we performed two sets of acquisitions: (i) one plane-by-plane acquisition for reference, and (ii) one compressed sensing acquisition. The plane-by-plane acquisition was performed by setting the measurement matrix $\mathbf{A}$ equal to the identity matrix in the acquisition software ($\mathbf{A} = \mathbf{I_N}$) (Fig. 1(b)), in order to facilitate comparisons with the compressed imaging scheme and acquiring a z-stack with 101 frames at a fixed exposure time of Δt =100 ms or 200 ms depending on the sample (see below), corresponding to a SNR of the reference image of ~ 10-15. For the compressed sensing acquisition, we used a Fourier measurement matrix, depicted in Fig. 1(d), with the appropriate scaling (see Section 2.3) to ensure that the light dose delivered to the sample for each camera exposure was equal to or lower than in the plane-by-plane acquisition. We acquired 50 frames $F_1^{comp}, \dots F_{50}^{comp}$ with the same exposure time Δt, i.e. corresponding to a compression ratio of 1:2. To analyze higher compression ratios, we simply considered subsets of these frames $F_1^{comp}, F_2^{comp} \dots F_M^{comp}$, with $M < 50$. The highest compression ratio, 1:50, was obtained by keeping only the two first compressed frames F_1^{comp} and F_2^{comp}. Varying the compression ratio in this manner allowed us to assess the minimum number of measurements required to obtain an acceptable reconstruction quality.

Optics EXPRESS

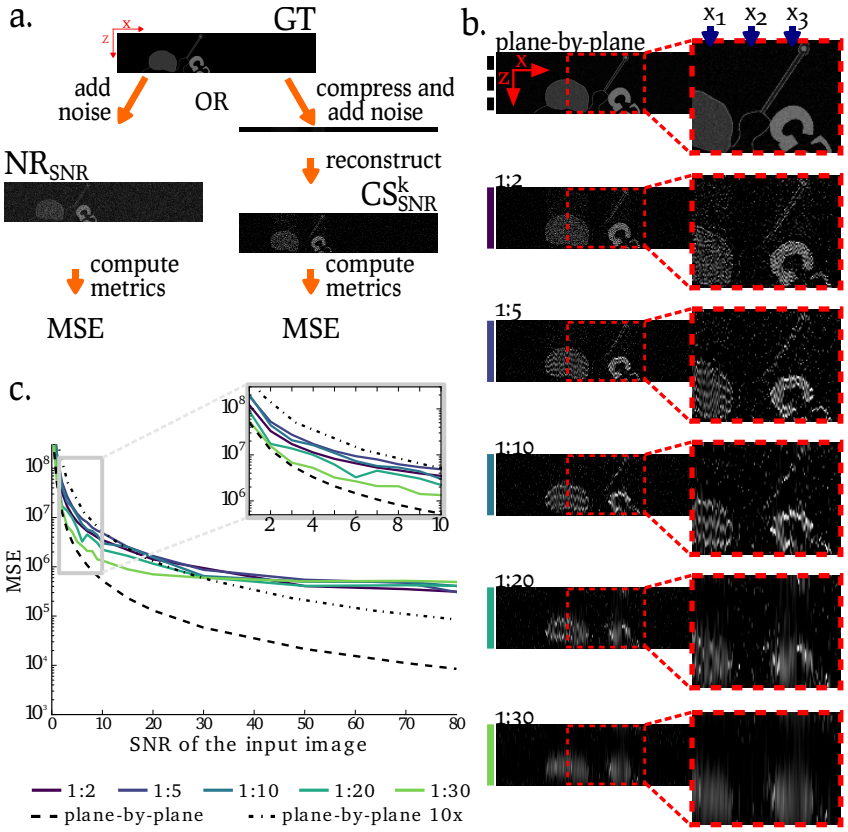


Fig. 2. Simulations comparing the compressed sensing scheme with the traditional plane-by-plane scheme under various compression ratios and SNR. **(a).** Principle of the simulation: from a generated ground truth image (**GT**) and a specified SNR, (left) a noisy reference ($\mathbf{NR_{SNR}}$) is generated by adding Poisson and Gaussian noise to the ground truth. In parallel (right), the ground truth is compressed (along the z axis) with a compression ratio κ and an equivalent amount of noise is added at the same time as the compression (see main text for details). The compressed images are further decompressed (image $\mathbf{CS^{\kappa}_{SNR}}$) and the mean square error (MSE) is computed with respect to the ground truth. **(b)**. Examples of simulated images in the (x, z) plane. From top to bottom: (plane-by-plane) **GT**, (1:2) to (1:20) $\mathbf{CS^{\kappa}_{SNR}}$ image reconstructed at a SNR of 20 and with a compression ratio of 1:2 to 1:20. The blue arrows represent lines of low sparsity, high sparsity and medium sparsity (respectively x_1, x_2 and x_3). **(c).** quality of the reconstruction (assessed by the MSE with respect to **GT**) for various SNR and compression-ratios. The dashed line is the MSE of the noisy reference $\mathbf{NR_{SNR}}$ with respect to the ground truth **GT**. The dash-dotted line is the MSE of the noisy reference acquired with a ten times lower exposure time $\mathbf{NR^{10x}_{SNR}}$. Inset: close-up of the low SNR region (SNR=1-10).

Optics EXPRESS

4.2. Results on fluorescent beads

We first imaged fluorescent beads (Tetraspeck 100 nm beads) on a glass coverslip illuminated with a 488 nm laser. The exposure time for a single frame was set to $\Delta t = 100$ ms and the stage was scanned over a 20 μm range (i.e. $\Delta z = 0.2\mu m$).

Figure 3(c) compares a 3D image of two fluorescent beads obtained through plane-by-plane imaging to 3D images reconstructed from the compressed acquisition for increasing compression ratios κ (from 1:2 to 1:50). It is apparent that qualitatively, the bead signal is reconstructed in the correct position for all compression ratios, including the highest (1:50), which corresponds to only 2 frames (instead of 101 in the plane-by-plane acquisition). As the compression ratio is increased from 1:2 to 1:50, the reconstructed images of the two beads (i.e. PSFs) progressively deteriorate and became significantly distorted for compression ratios $\geq$ 1:20. However, up to a compression ratio of 1:10, all the beads and the fine features of the PSF shape are properly and accurately reproduced, see Fig. 3(b). This is also apparent in the line profiles in Fig. 3(d). Thus, this experiment illustrates that high quality reconstruction of 3D images is possible in compressed imaging using 1s of total acquisition time, compared to 10s in the plane-by-plane imaging scheme, thereby representing a 10 fold speedup.

4.3. Results on fixed cells

Encouraged by these results on beads, we proceeded to imaging fixed cells. Mouse embryonic stem cells (mESCs) were seeded on glass coverslips and fixed in 4% paraformaldehyde (PFA). The cells were then stained for actin with a fluorescently labeled probe (phalloidin-RFP) and imaged in an oxygen-scavenging medium. The camera exposure time was set to $\Delta t = 200$ ms and the z scanning range to 20 μm ($\Delta z = 0.2\mu m$).

Results are shown in Fig. 4(a), where the 3D reconstruction of a sample imaged with 1:5 compression (right) is compared to the plane-by-plane acquisition (left). It is apparent that in both lateral (x, z) and axial (x, y) sections, fine and large details are successfully reconstructed. Although some high frequency details are lost in the (x, z) plane, most features are faithfully reproduced. This observation is further confirmed by the maximum intensity projection of the reconstructed stacks at various compression ratios, see Fig. 4(b). For compression ratios up to 1:5, the reconstructions show intensity profiles very similar to those of the plane-by-plane reference image. At higher compression ratios, however, significant artifacts become visible along the z axis (profiles in Fig. 4(a) and 4(b) bottom). Artifact-free reconstructions with higher compression ratios might be possible using a number of possible improvements (see Discussion). Nevertheless, these results already indicate that our compressed sensing approach is a viable method to achieve substantial reductions of acquisition time (and light exposure) for 3D imaging of biological samples.

5. Compressed epifluorescence microscopy

5.1. Implementation

We also implemented our compressed sensing scheme on a standard epifluorescence microscope, an almost ubiquitous instrument in cell biology labs. Since our imaging strategy relies only on the synchronization of the stage position and the light modulation, it is very versatile and suitable for a wide range of microscopes.

An epifluorescence microscope (Nikon Eclipse TI) equipped with an AOTF (AOTFnC-400-650-TN, AA Optoelectronics) and a z piezo stage (Nano-ZL 500, Mad City Labs) is controlled using an Arduino microcontroller (Genuino Uno) to synchronize the AOTF and the stage through their analog input, based on the camera fire signal, see Fig. 5(a).

The microscope was controlled using MicroManager [32], and custom firmware was written for the Arduino, allowing for a software switch between plane-by-plane and compressed sensing

Optics EXPRESS

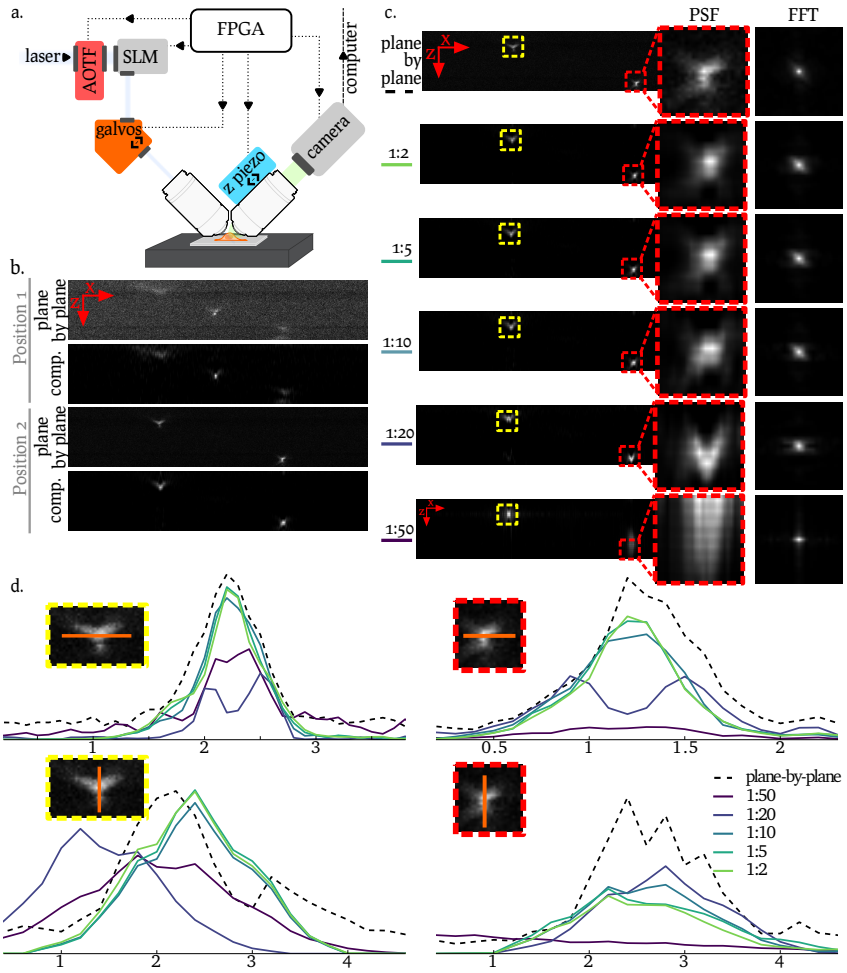


Fig. 3. Compressed imaging reconstructions of fluorescent beads acquired with a lattice light sheet microscope. **(a).** Principle of a lattice light sheet microscope: two objectives at a 90°angle are used to observe the sample. The light sheet is generated through a spatial light modulator (SLM) and associated optics. The focus is adjusted by a coordinated move of the z piezo (that translates the observation objective) and of the z galvo (that translates the light sheet). Synchronization is achieved by a FPGA (Field-Programmable Gate Array). **(b).** Sample reconstructions (*compressed*) in the (x, z) plane from a 1:10 compressed acquisition and the corresponding acquisition in the plane-by-plane imaging scheme for two y positions (termed position 1 and position 2). **(c).** Compressed imaging sample reconstructions at increasing compression ratios (from 1:2 to 1:50) compared to the plane-by-plane scheme (top). For each reconstruction, a close-up of the PSF is displayed (*PSF* column) and the 2D frequency content of the PSF is displayed next to the PSF (*FFT* column). The PSF shown in red is the typical response of the LLSM whereas the shape of the yellow PSF is likely due to a defect in the bead. **(d).** Line profile across the two highlighted PSFs (in yellow and red, respectively left and right). *top* Close-up along the x axis, *bottom* close-up along the z axis. The colors correspond to different compression ratios and the dotted line to the plane-by-plane reference. Horizontal axis in μm. A field of view in the (x, z) plane is 50x20 μm (512x101 px). A full 3D reconstructions is provided in blue)Visualization 1.

Optics EXPRESS

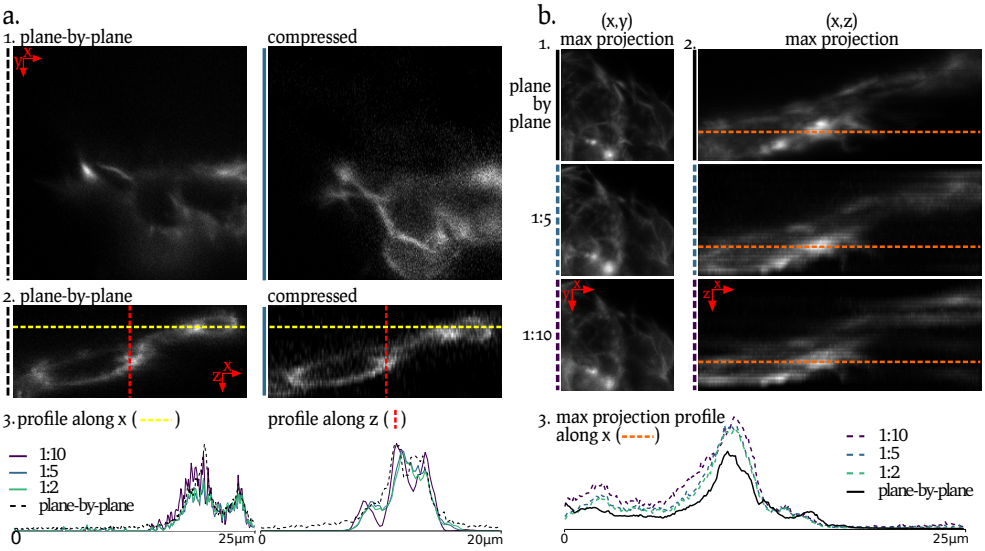


Fig. 4. Compressed imaging reconstruction of phalloidin-RFP-stained, fixed mouse embryonic stem cells (mESCs) acquired with a lattice light sheet microscope. **(a).** Example reconstruction from a 5 fold compression ratio (right) and the corresponding plane-by-plane reference acquisition (left) (*1., top* in the (x, y) plane and (*2., middle*) in the (x, z) plane. Dotted lines indicate the location of the line profiles presented in (panel *3., bottom*) (left) profile along the x axis (yellow dotted line of panel 2.) for increasing compression ratios. (right) profile along the z axis (red dotted line of panel 3.) the curves are the average over 3 planes in the y dimension. **(b).** Maximum intensity projection in the (x, y) plane (*1., left*) and in the (x, z) plane (*2., right*) of the plane-by-plane stack (top) and reconstructed stack at increasing compression ratios (two bottom pictures). The orange dotted line shows the location of the line profile displayed in panel *3., bottom*: line profile of the reconstruction at increasing compression ratios (dotted line) compare to the plane-by-plane reference (continuous line). A field of view in the (x, z) plane is 25x20 μm (256x101 px) and a field of view in the (x, y) plane is 25x25 μm (256x256 px). A full 3D reconstructions is provided in blue)Visualization 2.

acquisition. In practice, the measurement matrix $\mathbf{A}$ is first loaded to the Arduino, then the focus is adjusted in the plane-by-plane imaging mode and an acquisition sequence is set using a custom MicroManager plugin. Finally, the Arduino is switched to the compressed sensing mode and the images are acquired and handled using the usual MicroManager logic.

5.2. Results on fluorescent beads

We imaged fluorescent beads on a glass coverslip with an axial range of ~ 100 μm with a 60x/1.3 NA oil immersion objective, an exposure time of $\Delta t = 200$ ms illuminated with 561 nm laser, leading to an average SNR of 15. Results are shown in Fig. 5(b)–Fig. 5(d) for varying compression ratios κ. It is apparent from Fig. 5(c) that the reconstructed images qualitatively recover the beads position accurately for compression ratios up to 1:10. Interestingly, the reconstructed images display a decreased level of background noise, revealing the PSF shape of out-of-plane beads (i.e. beads in another $y = constant$ plane) that are not visible in the noisy, plane-by-plane imaging reference, see Fig. 4(b). Although the location of the PSF in the reconstructed image is consistent with the plane-by-plane reference depicted in Fig. 4(d), the PSF in the latter is significantly sharper than in the reconstructed image. This might be due either to an

Optics EXPRESS

insufficient synchronization of the AOTF with the stage or to an inaccurate PSF model used for the reconstruction. For higher compression ratios ($\geq$ 1:20), the PSF is no longer well localized in z.

These results indicate that our compressed sensing approach can be successfully applied to an epifluorescence setup and achieve roughly ten-fold compression ratios. Given the large availability of epifluorescence microscopes, this shows that the benefits of our compressed imaging approach can be made widely accessible with little effort.

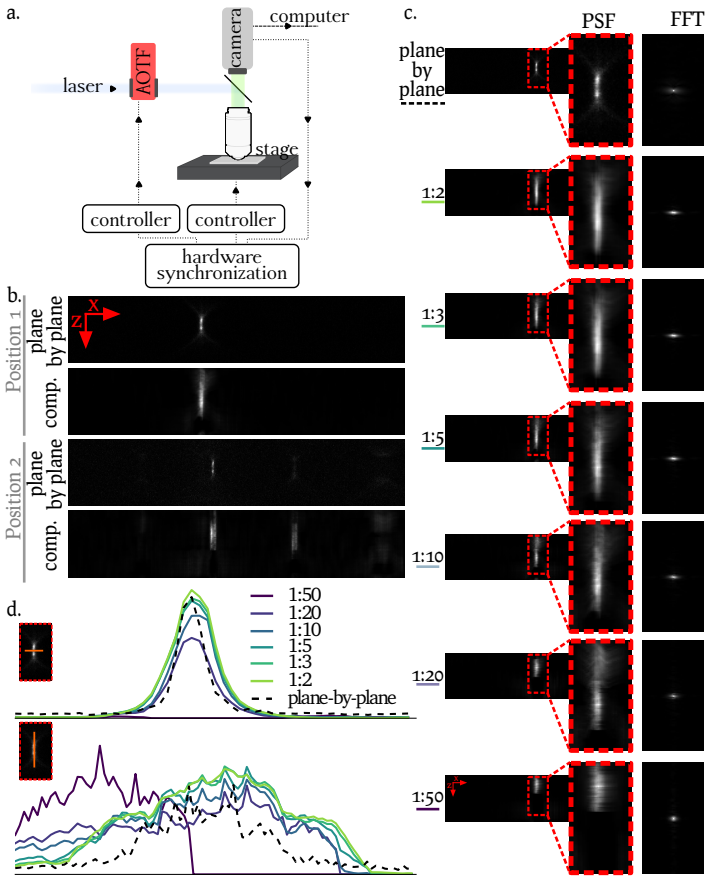


Fig. 5. Compressed imaging reconstructions of fluorescent beads acquired with an epifluorescence microscope. **(a).** Principle of the epifluorescence microscope: both the AOTF and the motorized stage in z are synchronized by hardware (Arduino). **(b).** Example reconstructions (*compressed*) in the (x, z) plane from a 1:10 compressed acquisition and the corresponding acquisition in the traditional imaging scheme (*plane-by-plane*) for two y positions (termed position 1 and position 2). **(c).** Sample reconstructions at increasing compression ratios (from 1:2 to 1:50) compared to the plane-by-plane imaging scheme (top). **(d).** x and z profiles of one selected PSF (highlighted in panel c) reconstructed from various compression ratios. The black dotted line represents the plane-by-plane reference. A full 3D reconstructions is provided in blue)Visualization 3.

Optics EXPRESS

6. Discussion

In this paper, we present a new compressed sensing scheme for 3D fluorescence microscopy that can be applied to a wide range of microscopes. We validate our method through simulations, and demonstrate its feasability on a lattice light sheet microscope and on an epifluorescence microscope. We achieve reductions in image acquisition time and light exposure (compression ratios) of up to ten fold on both setups.

Most previously proposed compressed sensing strategies for biological microscopy require non-trivial modifications to microscope hardware, such as adding a digital micromirror device to the light path [17], or conjugating the camera with the back pupil plane to perform Fourier space imaging [16]. By contrast, our compressed sensing scheme is adaptable to any type of epifluorescence microscope, as long as the user can control both the stage position and the illumination intensity. Our lattice light sheet microscope implementation required software adjustments due to its complex nature, while implementation on an epifluorescence microscope required only one extra microcontroller (to ensure proper synchronization of the camera, AOTF and stage). We provide full schematics of the Arduino setup, together with a MicroManager plugin, which allows to quickly switch between plane-by-plane imaging and the compressed sensing imaging scheme.

Since our approach modulates the light intensity before it reaches the sample it results in a reduced excitation light dose to the sample. This is of crucial importance, since fluorescence imaging causes damage to the sample, ranging from photobleaching of the fluorescent probes to various metabolic and developmental defects. Furthermore, the light dose delivered to the sample is inversely proportional to the compression ratio: a ten fold compression ratio yields a ten fold reduction of the light dose at the sample compared to plane-by-plane imaging. Since phototoxicity is a nonlinear effect [33], we expect our compressed imaging method to allow dramatic reductions in photodamage. We note that due to the loss of high-frequency information in the implementation described here, the current approach is best suited for applications where long-term or high-temporal resolution 4D observation of larger scale fluorescently labeled structures is required. Potential applications of our scheme include long term imaging of transient cell cycle events in a developing fly embryo, the propagation of calcium influx in beating cardiomyocytes, the observation of signaling events or motility in live organisms or in cells growing in 3D matrices.

Another major benefit of our compressed imaging scheme is an increase in the temporal resolution of 3D imaging, thus allowing faster acquisition at a given SNR. For example, if plane-by-plane imaging requires 200 planes with 10 ms exposure each, i.e. a total of 2 s to acquire a given 3D volume, a compressed imaging scheme with a compression ratio of 10 will require only 200 ms for the same volume. Our simulations in Fig. 2 demonstrate that this scheme yields images of lower MSE than a plane-by-plane z-stack acquired with a 1 ms exposure time (that is a 200 ms overall acquisition time). This reduction in acquisition time should enable an equivalent increase in the temporal resolution of dynamic 3D microscopy of living biological samples. We anticipate that future work will build on the proposed approach to explore the potential of axially compressed imaging for faster 3D live cell imaging.

In this context, several challenges and perspectives for improvement are worth mentioning. First, better piezo hardware and/or calibration could reduce the mismatch between the theoretical sensing matrix and its experimental counterpart. Second, as in plane-by-plane imaging, our current method assumes that the imaged structure remains immobile throughout compressed acquisition of the 3D volume, which is rarely true in living samples. While sample movements can result in a blurred image or duplicated objects with traditional plane-by-plane imaging, it remains to be explored how these movements may distort reconstructed images in our compressed sensing scheme. As previously shown in the MRI field, methods to address these problems can be developed [30]. Third, computational strategies to accelerate image processing will be important to efficiently analyze the thousands of images generated in dynamic 3D imaging [34, 35]. Finally,

Optics EXPRESS

our method currently assumes sparsity in the image domain along the optical axis only and used a Fourier sensing matrix. Ten fold compression ratios were achieved on images with moderate degrees of sparsity, for which this sensing matrix is not optimal. We therefore expect that larger compression ratios can be achieved using 3D sparsity models better adapted to the imaged structures, e.g. using dictionary learning, or optimized sensing matrices [23, 36].

7. Conclusion

In this work, we demonstrate a widely applicable 3D compressed sensing scheme in which images are compressed along the z axis during acquisition. This scheme can be implemented with little or no hardware modification on a wide range of microscopes. We first validated the feasability of our approach under noisy conditions using simulations, and then demonstrated the method experimentally on a lattice light sheet microscope and on an epifluorescence microscope, where we achieved roughly ten fold imaging speedup. This approach can be used to increase the temporal resolution or extend imaging time (through a reduction in light dose) in 3D fluorescence imaging applications.

Software and data availability

A Python port of SPIRALTAP [27] is available on GitHub https://github.com/imodpasteur/pySPIRALTAP (doi:10.5281/zenodo.439691. The datasets used in this work are available on the Zenodo repository (doi:10.5281/zenodo.439689) and the analysis scripts were deposited on GitHub https://github.com/imodpasteur/CompressedSensingMicroscopy3D (doi:10.5281/zenodo.439690). The MicroManager plugin and Arduino device adapter are available on https://github.com/imodpasteur/ArduinoCompressedSensing.

Funding

California Institute of Regenerative Medicine (CIRM) LA1-08013 to X.D.; National Institutes of Health (NIH) UO1-EB021236 to X.D.; National Institutes of Health (NIH) U54-DK107980 to X.D.; the Région Île de France (DIM Malinf) to C.Z; Visiting scholarship from the Siebel Stem Cell Foundation to C.Z; Institut Pasteur to C.Z.

Acknowledgments

We are very grateful to Zach Harmany who made available the SPIRALTAP software under an open source license and to Moran Mordechay for useful discussions on further optimization of the imaging scheme. We would like to thank the whole IMOD lab and Darzacq labs for insightful discussions and suggestions to this work. We thank Astou Tangara and the Betzig Lab (HHMI Janelia Research Campus) for their help in constructing the LLSM.

This work used the computational and storage services (TARS cluster) provided by the IT department at Institut Pasteur, Paris.

Part V

References

Bibliography

Abad, María et al. (2013). "Reprogramming in Vivo Produces Teratomas and iPS Cells with Totipotency Features". In: *Nature* 502.7471. 00138, pp. 340–345. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature12586. URL: http://www.nature.com/doifinder/10.1038/nature12586 (visited on 08/23/2016).

Akhtar, Waseem et al. (2013). "Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel". In: *Cell* 154.4. 00083, pp. 914–927. ISSN: 00928674. DOI: 10.1016/j.cell.2013.07.018. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867413008891 (visited on 02/07/2017).

Alabert, Constance et al. (2014). "Nascent Chromatin Capture Proteomics Determines Chromatin Dynamics during DNA Replication and Identifies Unknown Fork Components". In: *Nature Cell Biology* 16.3. 00094, pp. 281–293. ISSN: 1465-7392, 1476-4679. DOI: 10.1038/ncb2918. URL: http://www.nature.com/doifinder/10.1038/ncb2918 (visited on 04/01/2017).

Alexander, Jeffrey M. et al. (2018). "Live-Cell Imaging Reveals Enhancer-Dependent Sox2 Transcription in the Absence of Enhancer Proximity". In: *bioRxiv*, p. 409672. DOI: 10.1101/409672. URL: https://www.biorxiv.org/content/early/2018/09/05/409672 (visited on 09/06/2018).

Alipour, Elnaz and John F. Marko (2012). "Self-Organization of Domain Structures by DNA-Loop-Extruding Enzymes". In: *Nucleic Acids Research* 40.22, pp. 11202–11212. ISSN: 0305-1048. DOI: 10.1093/nar/gks925. URL: https://academic.oup.com/nar/article/40/22/11202/1155465 (visited on 04/06/2019).

Altmeyer, Matthias et al. (2015). "Liquid Demixing of Intrinsically Disordered Proteins Is Seeded by Poly(ADP-Ribose)". In: *Nature Communications* 6, p. 8088. ISSN: 2041-1723. DOI: 10.1038/ncomms9088. URL: https://www.nature.com/articles/ncomms9088 (visited on 02/01/2018).

Amati, Bruno et al. (1992). "Transcriptional Activation by the Human C-Myc Oncoprotein in Yeast Requires Interaction with Max". In: 00436.

Ambrose, E. J. (1956). "A Surface Contact Microscope for the Study of Cell Movements". In: *Nature* 178.4543, p. 1194. ISSN: 1476-4687. DOI: 10.1038/1781194a0. URL: https://www.nature.com/articles/1781194a0 (visited on 04/06/2019).

Amitai, Assaf (2018). "Chromatin Configuration Affects the Dynamics and Distribution of a Transiently Interacting Protein". In: *bioRxiv*, p. 246231. DOI: 10.1101/246231. URL: https://www.biorxiv.org/content/early/2018/01/10/246231 (visited on 01/11/2018).

Amitai, Assaf et al. (2017). "Visualization of Chromatin Decompaction and Break Site Extrusion as Predicted by Statistical Polymer Modeling of Single-Locus Trajectories". In: *Cell Reports* 18.5, pp. 1200–1214. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2017.01.018. URL: http://www.sciencedirect.com/science/article/pii/S2211124717300542 (visited on 03/02/2018).

Ando, Tadashi and Jeffrey Skolnick (2010). "Crowding and Hydrodynamic Interactions Likely Dominate in Vivo Macromolecular Motion". In: *Proceedings of the National Academy of Sciences* 107.43, pp. 18457–18462. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1011354107. pmid: 20937902. URL: http://www.pnas.org/content/107/43/18457 (visited on 10/29/2017).

Andresen, Cecilia et al. (2012). "Transient Structure and Dynamics in the Disordered C-Myc Transactivation Domain Affect Bin1 Binding". In: *Nucleic Acids Research* 40.13, pp. 6353–6366. ISSN: 0305-1048. DOI: 10.1093/nar/gks263. URL: https://academic.oup.com/nar/article/40/13/6353/1013958 (visited on 11/09/2017).

Andrews, J. Owen et al. (2017). "qSR: A Software for Quantitative Analysis of Single Molecule and Super-Resolution Data". In: *bioRxiv*. 00000. DOI: 10.1101/146241. URL: http://biorxiv.org/content/early/2017/06/05/146241.abstract.

Andronov, Leonid, Yves Lutz, et al. (2016). "SharpViSu: Integrated Analysis and Segmentation of Super-Resolution Microscopy Data". In: *Bioinformatics* 32.14. 00005, pp. 2239–2241. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btw123. URL: https://academic.oup.com/bioinformatics/article/32/14/2239/1742770/SharpViSu-integrated-analysis-and-segmentation-of (visited on 06/11/2017).

Andronov, Leonid, Jonathan Michalon, et al. (2017). "3D Clustering Analysis of Super-Resolution Microscopy Data by 3D Voronoi Tessellations". In: *bioRxiv*. 00000. DOI: 10.1101/146456. URL: http://biorxiv.org/content/early/2017/06/07/146456.abstract.

Andronov, Leonid, Igor Orlov, et al. (2016). "ClusterViSu, a Method for Clustering of Protein Complexes by Voronoi Tessellation in Super-Resolution Microscopy". In: *Scientific Reports* 6.1. 00014. ISSN: 2045-2322. DOI: 10.1038/srep24084. URL: http://www.nature.com/articles/srep24084 (visited on 06/11/2017).

Arbona, Jean-Michel et al. (2017). "Inferring the Physical Properties of Yeast Chromatin through Bayesian Analysis of Whole Nucleus Simulations". In: *Genome Biology* 18.1. 00000. ISSN: 1474-760X. DOI: 10.1186/s13059-017-1199-x. URL: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1199-x (visited on 05/19/2017).

Aristov, Andrey et al. (2018). "ZOLA-3D Allows Flexible 3D Localization Microscopy over an Adjustable Axial Range". In: *Nature Communications* 9.1, p. 2409. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04709-4. URL: https://www.nature.com/articles/s41467-018-04709-4 (visited on 06/19/2018).

Arya, Gaurav and Tamar Schlick (2006). "Role of Histone Tails in Chromatin Folding Revealed by a Mesoscopic Oligonucleosome Model". In: *Proceedings of the National Academy of Sciences* 103.44, pp. 16236–16241. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0604817103. pmid: 17060627. URL: http://www.pnas.org/content/103/44/16236 (visited on 08/29/2018).

Auduge, Nicolas et al. (2018). "Chromatin Condensation Fluctuations Rather Than Steady-State Predict Chromatin Accessibility". In: DOI: 10.1101/365700. URL: http://biorxiv.org/lookup/doi/10.1101/365700 (visited on 09/07/2018).

Avcu, Neslihan and Nacho Molina (2016). "Chromatin Structure Shapes the Search Process of Transcription Factors". In: *bioRxiv*, p. 050146. DOI: 10.1101/050146. URL: https://www.biorxiv.org/content/early/2016/04/25/050146 (visited on 06/12/2018).

Avivi, Shira et al. (2017). "Visualizing Nuclear RNAi Activity in Single Living Human Cells". In: *Proceedings of the National Academy of Sciences* 114.42, E8837–E8846. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1707440114. URL: http://www.pnas.org/content/114/42/E8837 (visited on 10/17/2017).

Baddeley, David, Mark B. Cannell, and Christian Soeller (2010). "Visualization of Localization Microscopy Data". In: *Microscopy and Microanalysis* 16.01, pp. 64–72. ISSN: 1431-9276, 1435-8115. DOI: 10.1017/S143192760999122X. URL: http://www.journals.cambridge.org/abstract_S143192760999122X (visited on 11/18/2017).

Bai, Xiaoyu and Peter G. Wolynes (2015). "On the Hydrodynamics of Swimming Enzymes". In: *The Journal of Chemical Physics* 143.16, p. 165101. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.4933424. URL: http://aip.scitation.org/doi/10.1063/1.4933424 (visited on 04/25/2019).

Bak, A L, J Zeuthen, and F H Crick (1977). "Higher-Order Structure of Human Mitotic Chromosomes." In: *Proceedings of the National Academy of Sciences of the United States of America* 74.4, pp. 1595–1599. ISSN: 0027-8424. pmid: 266199. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC430837/ (visited on 04/06/2019).

Baldi, Sandro et al. (2018). "Genome-Wide Measurement of Local Nucleosome Array Regularity and Spacing by Nanopore Sequencing". In: *bioRxiv*, p. 272526. DOI: 10.1101/272526. URL: https://www.biorxiv.org/content/early/2018/03/07/272526 (visited on 03/09/2018).

Balleza, Enrique, J. Mark Kim, and Philippe Cluzel (2017). "Systematic Characterization of Maturation Time of Fluorescent Proteins in Living Cells". In: *Nature Methods*. ISSN: 1548-7105. DOI: 10.1038/nmeth.4509. URL: https://www.nature.com/articles/nmeth.4509 (visited on 12/06/2017).

Balzarotti, Francisco et al. (2017). "Nanometer Resolution Imaging and Tracking of Fluorescent Molecules with Minimal Photon Fluxes". In: *Science* 355.6325. 00002, pp. 606–612. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aak9913. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aak9913 (visited on 03/15/2017).

Banaz, Nehir, Jarno Makela, and Stephan Uphoff (2018). "Choosing the Right Label for Single-Molecule Tracking in Live Bacteria: Side-by-Side Comparison of Photoactivatable Fluorescent Protein and Halo Tag Dyes". In: DOI: 10.1101/381665. URL: http://biorxiv.org/lookup/doi/10.1101/381665 (visited on 08/25/2018).

Bancaud, Aurélien et al. (2009). "Molecular Crowding Affects Diffusion and Binding of Nuclear Proteins in Heterochromatin and Reveals the Fractal Organization of Chromatin". In: *The EMBO Journal* 28.24, pp. 3785–3798. ISSN: 0261-4189, 1460-2075. DOI: 10.1038/emboj.2009.340. URL: http://emboj.embopress.org/cgi/doi/10.1038/emboj.2009.340 (visited on 09/25/2017).

Bannister, Andrew J and Tony Kouzarides (2011). "Regulation of Chromatin by Histone Modifications". In: *Cell Research* 21.3, pp. 381–395. ISSN: 1001-0602, 1748-7838. DOI: 10.1038/cr.2011.22. URL: http://www.nature.com/articles/cr201122 (visited on 04/12/2018).

Baptista, Tiago et al. (2017). "SAGA Is a General Cofactor for RNA Polymerase II Transcription". In: *Molecular Cell*. ISSN: 10972765. DOI: 10.1016/j.molcel.2017.08.016. URL: http://linkinghub.elsevier.com/retrieve/pii/S1097276517306135 (visited on 10/06/2017).

Barbieri, Letizia, Enrico Luchinat, and Lucia Banci (2016). "Characterization of Proteins by In-Cell NMR Spectroscopy in Cultured Mammalian Cells". In: *Nature Protocols* 11.6, pp. 1101–1111. ISSN: 1754-2189. DOI: 10.1038/nprot.2016.061. URL: http://www.nature.com/nprot/journal/v11/n6/full/nprot.2016.061.html (visited on 10/30/2017).

Bardales, Jorge A et al. (2018). "Selective Activation of Alternative MYC Core Promoters by Wnt-Responsive Enhancers". In: p. 15.

Barnes, Stephanie L et al. (2018). "Mapping DNA Sequence to Transcription Factor Binding Energy in Vivo". In: DOI: 10.1101/331124. URL: http://biorxiv.org/lookup/doi/10.1101/331124 (visited on 07/02/2018).

Basset, Antoine et al. (2015). "Adaptive Spot Detection With Optimal Scale Selection in Fluorescence Microscopy Images". In: *IEEE Transactions on Image Processing* 24.11, pp. 4512–4527. ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2015.2450996. URL: http://ieeexplore.ieee.org/document/7140815/ (visited on 02/21/2018).

Basu, Srinjan et al. (2018). "FRET-Enhanced Photostability Allows Improved Single-Molecule Tracking of Proteins and Protein Complexes in Live Mammalian Cells". In: *Nature Communications* 9.1, p. 2520. ISSN: 2041-1723. DOI:

10.1038/s41467-018-04486-0. URL: http://www.nature.com/articles/s41467-018-04486-0 (visited on 04/25/2019).

Batsche, Eric and Chantal Cremisi (1999). "Opposite Transcriptional Activity between the Wild Type C-Myc Gene Coding for c-Myc1 and c-Myc2 Proteins and c-Myc1 and c-Myc2 Separately". In: *Oncogene* 18.41. 00018, pp. 5662–5671. URL: https://www.researchgate.net/profile/Eric_Batsche/publication/12776398_Opposite_transcriptional_activity_between_the_wild_type_c-myc_gene_coding_for_c-Myc1_and_c-Myc2_proteins_and_c-Myc1_and_c-Myc2_separately/links/0fcfd50ad35005e636000000.pdf (visited on 08/23/2016).

Bayliss, Richard et al. (2017). "A Moving Target: Structure and Disorder in Pursuit of Myc Inhibitors". In: *Biochemical Society Transactions* 45.3. 00000, pp. 709–717. ISSN: 0300-5127, 1470-8752. DOI: 10.1042/BST20160328. URL: http://biochemsoctrans.org/lookup/doi/10.1042/BST20160328 (visited on 07/19/2017).

Bazett-Jones, David P. et al. (2008). "Elucidating Chromatin and Nuclear Domain Architecture with Electron Spectroscopic Imaging". In: *Chromosome Research* 16.3, pp. 397–412. ISSN: 0967-3849, 1573-6849. DOI: 10.1007/s10577-008-1237-3. URL: http://link.springer.com/10.1007/s10577-008-1237-3 (visited on 12/05/2017).

Beagrie, Robert A. et al. (2017). "Complex Multi-Enhancer Contacts Captured by Genome Architecture Mapping". In: *Nature*. 00000. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature21411. URL: http://www.nature.com/doifinder/10.1038/nature21411 (visited on 03/15/2017).

Beck, Samuel et al. (2014). "CpG Island-Mediated Global Gene Regulatory Modes in Mouse Embryonic Stem Cells". In: *Nature Communications* 5. 00008, p. 5490. ISSN: 2041-1723. DOI: 10.1038/ncomms6490. URL: http://www.nature.com/doifinder/10.1038/ncomms6490 (visited on 07/17/2017).

Bedino, James H (2003). "Embalming Chemistry: Glutaraldehyde versus Formaldehyde". In: p. 19.

Beliveau, Brian J. et al. (2015). "Single-Molecule Super-Resolution Imaging of Chromosomes and *in Situ* Haplotype Visualization Using Oligopaint FISH Probes". In: *Nature Communications* 6, p. 7147. ISSN: 2041-1723. DOI: 10.1038/ncomms8147. URL: https://www.nature.com/articles/ncomms8147 (visited on 04/06/2019).

Bell, Jason C., David Jukam, Nicole A. Teran, Viviana I. Risca, Owen K. Smith, Whitney L. Johnson, Jan M. Skotheim, et al. (2018). "Chromatin-Associated RNA Sequencing (ChAR-Seq) Maps Genome-Wide RNA-to-DNA Contacts". In: *eLife* 7, e27024. ISSN: 2050-084X. DOI: 10.7554/eLife.27024. URL: https://elifesciences.org/articles/27024 (visited on 04/13/2018).

Bell, Jason C., David Jukam, Nicole A. Teran, Viviana I. Risca, Owen K. Smith, Whitney L. Johnson, Jan Skotheim, et al. (2017). "Chromatin-Associated RNA Sequencing (ChAR-Seq) Maps Genome-Wide RNA-to-DNA Contacts". In: *bioRxiv*, p. 118786. DOI: 10.1101/118786. URL: https://www.biorxiv.org/content/early/2017/03/20/118786 (visited on 10/02/2017).

Bello-Fernandez, C., G. Packham, and J. L. Cleveland (1993). "The Ornithine Decarboxylase Gene Is a Transcriptional Target of C-Myc." In: *Proceedings of the National Academy of Sciences* 90.16, pp. 7804–7808. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.90.16.7804. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.90.16.7804 (visited on 11/20/2018).

Belmont, Andrew S (2014). "Large-Scale Chromatin Organization: The Good, the Surprising, and the Still Perplexing". In: *Current Opinion in Cell Biology* 26. 00068, pp. 69–78. ISSN: 09550674. DOI: 10.1016/j.ceb.2013.10.002. URL: http://linkinghub.elsevier.com/retrieve/pii/S0955067413001555 (visited on 03/16/2017).

Benabdallah, Nezha S. et al. (2017). "PARP Mediated Chromatin Unfolding Is Coupled to Long-Range Enhancer Activation". In: *BioRxiv*. DOI: 10.1101/155325. URL: http://biorxiv.org/lookup/doi/10.1101/155325 (visited on 09/19/2018).

Ben-Avraham, Daniel and Shlomo Havlin (2000). *Diffusion and Reactions in Fractals and Disordered Systems*. 01061. Cambridge ; New York: Cambridge University Press. 316 pp. ISBN: 978-0-521-62278-3.

Benetatos, Leonidas, George Vartholomatos, and Eleftheria Hatzimichael (2014). "Polycomb Group Proteins and MYC: The Cancer Connection". In: *Cellular and Molecular Life Sciences* 71.2. 00024, pp. 257–269. ISSN: 1420-682X, 1420-9071. DOI: 10.1007/s00018-013-1426-x. URL: http://link.springer.com/10.1007/s00018-013-1426-x (visited on 08/23/2016).

Benham, Tim et al. (2015). "CEoptim: Cross-Entropy R Package for Optimization". In: *arXiv preprint arXiv:1503.01842*.

Bénichou, O. et al. (2010). "Geometry-Controlled Kinetics". In: *Nature Chemistry* 2.6. 00147, pp. 472–477. ISSN: 1755-4330, 1755-4349. DOI: 10.1038/nchem.622. URL: http://www.nature.com/doifinder/10.1038/nchem.622 (visited on 09/29/2016).

Bensaude, Olivier (2011). "Inhibiting Eukaryotic Transcription. Which Compound to Choose? How to Evaluate Its Activity?" In: *Transcription* 2.3, pp. 103–108. ISSN: 2154-1264, 2154-1272. DOI: 10.4161/trns.2.3.16172. URL: http://www.tandfonline.com/doi/abs/10.4161/trns.2.3.16172 (visited on 11/23/2017).

Bentley, D. L. and M. Groudine (1986). "Novel Promoter Upstream of the Human C-Myc Gene and Regulation of c-Myc Expression in B-Cell Lymphomas." In: *Molecular and Cellular Biology* 6.10, pp. 3481–3489. ISSN: 0270-7306, 1098-5549. DOI: 10.1128/MCB.6.10.3481. pmid: 3540591. URL: https://mcb.asm.org/content/6/10/3481 (visited on 03/30/2019).

Berdougo, Eli, Marie-Emilie Terret, and Prasad V. Jallepalli (2009). "Functional Dissection of Mitotic Regulators Through Gene Targeting in Human Somatic Cells". In: *Mitosis: Methods and Protocols*. Ed. by Andrew D. McAinsh. 00029. Totowa, NJ: Humana Press, pp. 21–37. ISBN: 978-1-60327-993-2. DOI: 10.1007/978-1-60327-993-2_2. URL: http://dx.doi.org/10.1007/978-1-60327-993-2_2.

Berger, Axel B. et al. (2008). "High-Resolution Statistical Mapping Reveals Gene Territories in Live Yeast". In: *Nature Methods* 5.12, pp. 1031–1037. ISSN: 1548-7105. DOI: 10.1038/nmeth.1266. URL: https://www.nature.com/articles/nmeth.1266 (visited on 04/06/2019).

Bergeron-Sandoval, Louis-Philippe, Nozhat Safaee, and Stephen W. Michnick (2016). "Mechanisms and Consequences of Macromolecular Phase Separation". In: *Cell* 165.5, pp. 1067–1079. ISSN: 0092-8674. DOI: 10.1016/j.cell.2016.05.026. URL: http://www.sciencedirect.com/science/article/pii/S0092867416305748 (visited on 02/01/2018).

Berglund, Andrew J. (2010). "Statistics of Camera-Based Single-Particle Tracking". In: *Physical Review E* 82.1. 00071. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.82.011917. URL: http://link.aps.org/doi/10.1103/PhysRevE.82.011917 (visited on 11/10/2016).

Bernard, Ora et al. (1983). "Sequence of the Murine and Human Cellular Myc Oncogenes and Two Modes of Myc Transcription Resulting from Chromosome Translocation in B Lymphoid Tumours." In: *The EMBO journal* 2.12. 00379, p. 2375. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC555460/ (visited on 04/03/2017).

Bernhem, Kristoffer and Hjalmar Brismar (2017). "SMLocalizer, a GPU Accelerated ImageJ Plugin for Single Molecule Localization Microscopy". In: *Bioinformatics*. DOI: 10.1093/bioinformatics/btx553. URL: https://academic.oup.com/bioinformatics/article/doi/10.1093/bioinformatics/btx553/4103397/SMLocalizer-a-GPU-accelerated-ImageJ-plugin-for (visited on 09/05/2017).

Beroukhim, Rameen et al. (2010). "The Landscape of Somatic Copy-Number Alteration across Human Cancers". In: *Nature* 463.7283, pp. 899–905. ISSN: 1476-4687. DOI: 10.1038/nature08822. URL: https://www.nature.com/articles/nature08822 (visited on 11/28/2018).

Berry, Hugues (2002). "Monte Carlo Simulations of Enzyme Reactions in Two Dimensions: Fractal Kinetics and Spatial Segregation". In: *Biophysical Journal* 83.4, pp. 1891–1901. ISSN: 00063495. DOI: 10.1016/S0006-3495(02)73953-2. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349502739532 (visited on 11/22/2017).

Betzig, E. et al. (2006). "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution". In: *Science* 313. 04715. DOI: 10.1126/science.1127344. URL: https://doi.org/10.1126/science.1127344.

Bhatia, Harsh et al. (2016). "Interactive Exploration of Atomic Trajectories through Relative-Angle Distribution and Associated Uncertainties". In: *Pacific Visualization Symposium (PacificVis), 2016 IEEE*. 00001. IEEE, pp. 120–127. URL: http://ieeexplore.ieee.org/abstract/document/7465259/ (visited on 04/02/2017).

Bianco, Simona et al. (2017). "Predicting Chromatin Architecture from Models of Polymer Physics". In: *Chromosome Research* 25.1, pp. 25–34. ISSN: 1573-6849. DOI: 10.1007/s10577-016-9545-5. URL: https://doi.org/10.1007/s10577-016-9545-5 (visited on 04/06/2019).

Bickel, Thomas (2007). "A Note on Confined Diffusion". In: *Physica A: Statistical Mechanics and its Applications* 377.1, pp. 24–32. ISSN: 03784371. DOI: 10.1016/j.physa.2006.11.008. arXiv: cond-mat/0604133. URL: http://arxiv.org/abs/cond-mat/0604133 (visited on 04/16/2019).

Bintu, Bogdan et al. (2018). "Super-Resolution Chromatin Tracing Reveals Domains and Cooperative Interactions in Single Cells". In: *Science* 362.6413, eaau1783. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aau1783. pmid: 30361340. URL: http://science.sciencemag.org/content/362/6413/eaau1783 (visited on 10/30/2018).

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe (2016). "Variational Inference: A Review for Statisticians". In: *arXiv preprint arXiv:1601.00670*. 00025. URL: http://arxiv.org/abs/1601.00670 (visited on 01/11/2017).

Boehning, Marc, Claire Dugast-Darzacq, Marija Rankovic, Anders S. Hansen, Tae-Kyung Yu, et al. (2018). "RNA Polymerase II Clustering through CTD Phase Separation". In: *bioRxiv*, p. 316372. DOI: 10.1101/316372. URL: https://www.biorxiv.org/content/early/2018/05/07/316372 (visited on 05/11/2018).

Boehning, Marc, Claire Dugast-Darzacq, Marija Rankovic, Anders S. Hansen, Taekyung Yu, et al. (2018). "RNA Polymerase II Clustering through Carboxy-Terminal Domain Phase Separation". In: *Nature Structural & Molecular Biology*. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/s41594-018-0112-y. URL: http://www.nature.com/articles/s41594-018-0112-y (visited on 08/26/2018).

Boettiger, Alistair N. et al. (2016). "Super-Resolution Imaging Reveals Distinct Chromatin Folding for Different Epigenetic States". In: *Nature* 529.7586, pp. 418–422. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature16496. URL: http://www.nature.com/doifinder/10.1038/nature16496 (visited on 03/24/2016).

Boija, Ann et al. (2018). "Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains". In: *Cell*. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2018.10.042. URL: https://www.cell.com/cell/abstract/S0092-8674(18)31398-9 (visited on 11/19/2018).

Boisvert, François-Michel et al. (2007). "The Multifunctional Nucleolus". In: *Nature Reviews Molecular Cell Biology* 8.7, pp. 574–585. ISSN: 1471-0080. DOI: 10.1038/nrm2184. URL: https://www.nature.com/articles/nrm2184 (visited on 04/06/2019).

Bolte, S. and F. P. Cordelieres (2006). "A Guided Tour into Subcellular Colocalization Analysis in Light Microscopy". In: *Journal of microscopy* 224.3, pp. 213–232.

Bolzer, Andreas et al. (2005). "Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes". In: *PLOS Biology* 3.5, e157. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0030157. URL: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0030157 (visited on 04/06/2019).

Bonev, Boyan et al. (2017). "Multiscale 3D Genome Rewiring during Mouse Neural Development". In: *Cell* 171.3, 557–572.e24. ISSN: 00928674. DOI: 10.1016/j.cell.2017.09.043. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867417311376 (visited on 11/11/2017).

Borcherds, Wade et al. (2014). "Disorder and Residual Helicity Alter P53-Mdm2 Binding Affinity and Signaling in Cells". In: *Nature Chemical Biology* 10.12, pp. 1000–1002. ISSN: 1552-4450, 1552-4469. DOI: 10.1038/nchembio.1668. URL: http://www.nature.com/doifinder/10.1038/nchembio.1668 (visited on 10/30/2017).

Börsch, Michael et al. (1998). "Conformational Changes of the H $^+$ -ATPase from *Escherichia Coli* upon Nucleotide Binding Detected by Single Molecule Fluorescence". In: *FEBS Letters* 437.3, pp. 251–254. ISSN: 00145793. DOI: 10.1016/S0014-5793(98)01247-2. URL: http://doi.wiley.com/10.1016/S0014-5793(98)01247-2 (visited on 04/25/2019).

Bosisio, Daniela et al. (2006). "A Hyper-dynamic Equilibrium between Promoter-bound and Nucleoplasmic Dimers Controls NF-κB-dependent Gene Activity". In: *The EMBO Journal* 25.4, pp. 798–810. ISSN: 0261-4189, 1460-2075. DOI: 10.1038/sj.emboj.7600977. pmid: 16467852. URL: http://emboj.embopress.org/content/25/4/798 (visited on 02/05/2018).

Botev, Zdravko I. et al. (2013). "The Cross-Entropy Method for Optimization". In: *Handbook of Statistics*. Vol. 31. Elsevier, pp. 35–59.

Bouchet, D. et al. (2019). "Probing Near-Field Light–Matter Interactions with Single-Molecule Lifetime Imaging". In: *Optica* 6.2, p. 135. ISSN: 2334-2536. DOI: 10.1364/OPTICA.6.000135. URL: https://www.osapublishing.org/abstract.cfm?URI=optica-6-2-135 (visited on 01/31/2019).

Bouzigues, Cédric and Maxime Dahan (2007). "Transient Directed Motions of GABAA Receptors in Growth Cones Detected by a Speed Correlation Index". In: *Biophysical Journal* 92.2, pp. 654–660. ISSN: 00063495. DOI: 10.1529/biophysj.106.094524. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349507708664 (visited on 11/18/2017).

Boyd, Nicholas et al. (2018). "DeepLoco: Fast 3D Localization Microscopy Using Neural Networks". In: DOI: 10.1101/267096. URL: http://biorxiv.org/lookup/doi/10.1101/267096 (visited on 04/09/2018).

Boyer, Denis, David S. Dean, Carlos Mejia-Monasterio, et al. (2013). "On Ergodic Least-Squares Estimators of the Generalized Diffusion Coefficient for Fractional Brownian Motion". In: *Physical Review E* 87.3. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.87.030103. arXiv: 1301.7638. URL: http://arxiv.org/abs/1301.7638 (visited on 08/02/2018).

Boyer, Denis, David S. Dean, Carlos Mejía-Monasterio, et al. (2012). "Optimal Estimates of the Diffusion Coefficient of a Single Brownian Trajectory". In: *Physical Review E* 85.3. 00027, p. 031136. URL: http://journals.aps.org/pre/abstract/10.1103/PhysRevE.85.031136 (visited on 11/10/2016).

Brady, Jacob P. et al. (2017). "Structural and Hydrodynamic Properties of an Intrinsically Disordered Region of a Germ Cell-Specific Protein on Phase Separation". In: *Proceedings of the National Academy of Sciences* 114.39, E8194–E8203. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1706197114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1706197114 (visited on 11/22/2017).

Branco, Miguel R. and Ana Pombo (2006). "Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations". In: *PLOS Biology* 4.5, e138. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0040138. URL: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040138 (visited on 06/12/2018).

Briane, Vincent, Myriam Vimond, and Charles Kervrann (2018). "An Overview of Diffusion Models for Intracellular Dynamics Analysis". working paper or preprint. URL: https://hal.inria.fr/hal-01966825.

Bronstein, I. et al. (2009). "Transient Anomalous Diffusion of Telomeres in the Nucleus of Mammalian Cells". In: *Physical Review Letters* 103.1. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.103.018102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.018102 (visited on 12/05/2017).

Büchel, Gabriele et al. (2017). "Association with Aurora-A Controls N-MYC-Dependent Promoter Escape and Pause Release of RNA Polymerase II during the Cell Cycle". In: *Cell Reports* 21.12, pp. 3483–3497. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2017.11.090. URL: http://www.sciencedirect.com/science/article/pii/S2211124717317576 (visited on 05/28/2018).

Burnecki, Krzysztof et al. (2012). "Universal Algorithm for Identification of Fractional Brownian Motion. A Case of Telomere Subdiffusion". In: *Biophysical Journal* 103.9, pp. 1839–1847. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.09.040. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349512011022 (visited on 09/25/2017).

Burov, S. et al. (2013). "Distribution of Directional Change as a Signature of Complex Dynamics". In: *Proceedings of the National Academy of Sciences* 110.49. 00034, pp. 19689–19694. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1319473110. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1319473110 (visited on 06/22/2017).

Busslinger, Georg A. et al. (2017). "Cohesin Is Positioned in Mammalian Genomes by Transcription, CTCF and Wapl". In: *Nature* 544.7651. 00000, pp. 503–507. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature22063. URL: http://www.nature.com/doifinder/10.1038/nature22063 (visited on 04/27/2017).

Butkevich, Alexey N. et al. (2017). "Hydroxylated Fluorescent Dyes for Live-Cell Labeling: Synthesis, Spectra and Super-Resolution STED". In: *Chemistry – A European Journal* 23.50, pp. 12114–12119. ISSN: 1521-3765. DOI: 10.1002/chem.201701216. URL: http://onlinelibrary.wiley.com/doi/10.1002/chem.201701216/abstract (visited on 09/21/2017).

Caccianini, Laura et al. (2015). "Single Molecule Study of Non-Specific Binding Kinetics of LacI in Mammalian Cells". In: *Faraday Discuss.* 184, pp. 393–400. ISSN: 1359-6640, 1364-5498. DOI: 10.1039/C5FD00112A. URL: http://xlink.rsc.org/?DOI=C5FD00112A (visited on 09/29/2017).

Cahoon, Cori K. et al. (2017). "Superresolution Expansion Microscopy Reveals the Three-Dimensional Organization of the *Drosophila* Synaptonemal Complex". In: *Proceedings of the National Academy of Sciences* 114.33. 00000, E6857–E6866. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1705623114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1705623114 (visited on 08/15/2017).

Callegari, Andrea et al. (2018). "Single-Molecule Dynamics and Genome-Wide Transcriptomics Reveal That NF-kB (P65)-DNA Binding Times Can Be Decoupled from Transcriptional Activation". In: *bioRxiv*, p. 255380. DOI: 10.1101/255380. URL: https://www.biorxiv.org/content/early/2018/01/28/255380 (visited on 02/05/2018).

Cambridge, Sidney B et al. (2009). "Doxycycline-Dependent Photoactivated Gene Expression in Eukaryotic Systems". In: *Nature Methods* 6.7. 00058, pp. 527–531. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1340. URL: http://www.nature.com/doifinder/10.1038/nmeth.1340 (visited on 08/23/2016).

Carey, Michael F., Craig L. Peterson, and Stephen T. Smale (2009). "Chromatin Immunoprecipitation (ChIP)". In: *Cold Spring Harbor Protocols* 2009.9, pdb.prot5279. ISSN: 1940-3402, 1559-6095. DOI: 10.1101/pdb.prot5279. pmid: 20147264. URL: http://cshprotocols.cshlp.org/content/2009/9/pdb.prot5279 (visited on 09/21/2018).

Caridi, Christopher P. et al. (2018). "Nuclear F-Actin and Myosins Drive Relocalization of Heterochromatic Breaks". In: *Nature* 559.7712, pp. 54–60. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-018-0242-8. URL: http://www.nature.com/articles/s41586-018-0242-8 (visited on 07/20/2018).

Carlini, Lina et al. (2015). "Correction of a Depth-Dependent Lateral Distortion in 3D Super-Resolution Imaging". In: *PloS one* 10.11. 00002, e0142949. URL: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0142949 (visited on 12/02/2016).

Carslaw, H. S. and J. C. Jaeger (1986). *Conduction of Heat in Solids*. 2nd ed. Oxford [Oxfordshire] : New York: Clarendon Press ; Oxford University Press. 510 pp. ISBN: 978-0-19-853368-9.

Carter, Peggy Sue, Marta Jarquin-Pardo, and Arrigo De Benedetti (1999). "Differential Expression of Myc1 and Myc2 Isoforms in Cells Transformed by eIF4E: Evidence for Internal Ribosome Repositioning in the Human c-Myc 5′UTR". In: *Oncogene* 18.30. 00053, pp. 4326–4335. ISSN: 09509232. DOI: 10.1038/sj.onc.1202890. URL: http://www.nature.com/doifinder/10.1038/sj.onc.1202890 (visited on 04/20/2017).

Cartwright, P. (2005). "LIF/STAT3 Controls ES Cell Self-Renewal and Pluripotency by a Myc-Dependent Mechanism". In: *Development* 132.5. 00731, pp. 885–896. ISSN: 0950-1991, 1477-9129. DOI: 10.1242/dev.01670. URL: http://dev.biologists.org/cgi/doi/10.1242/dev.01670 (visited on 08/14/2017).

Cattoglio, Claudia et al. (2018). "Architectural Features of 3D Genome Organization Revealed by Counting CTCF and Cohesin Molecules".

Chen, Baohui et al. (2013). "Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System". In: *Cell* 155.7. 00350, pp. 1479–1491. ISSN: 00928674. DOI: 10.1016/j.cell.2013.12.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867413015316 (visited on 08/23/2016).

Chen, B.-C. et al. (2014). "Lattice Light-Sheet Microscopy: Imaging Molecules to Embryos at High Spatiotemporal Resolution". In: *Science* 346.6208. 00184, pp. 1257998–1257998. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1257998. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1257998 (visited on 07/13/2016).

Chen, Hongtao et al. (2018). "Dynamic Interplay between Enhancer–Promoter Topology and Gene Activity". In: *Nature Genetics*, p. 1. ISSN: 1546-1718. DOI: 10.1038/s41588-018-0175-z. URL: https://www.nature.com/articles/s41588-018-0175-z (visited on 08/13/2018).

Chen, Hui, Hudan Liu, and Guoliang Qing (2018). "Targeting Oncogenic Myc as a Strategy for Cancer Treatment". In: *Signal Transduction and Targeted Therapy* 3.1, p. 5. ISSN: 2059-3635. DOI: 10.1038/s41392-018-0008-7. URL: https://www.nature.com/articles/s41392-018-0008-7 (visited on 03/30/2019).

Chen, Jiji et al. (2014). "Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells". In: *Cell* 156.6. 00144, pp. 1274–1285. ISSN: 00928674. DOI: 10.1016/j.cell.2014.01.062. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867414001974 (visited on 03/16/2017).

Chen, Kaifu et al. (2016). "The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses". In: *Molecular and Cellular Biology* 36.5, pp. 662–667. ISSN: 0270-7306. DOI: 10.1128/MCB.00970-14. pmid: 26711261. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4760223/ (visited on 10/17/2018).

Chen, Weizhong et al. (2018). "RNAs as Proximity-Labeling Media for Identifying Nuclear Speckle Positions Relative to the Genome". In: *iScience* 4, pp. 204–215. ISSN: 2589-0042. DOI: 10.1016/j.isci.2018.06.005. URL: http://www.sciencedirect.com/science/article/pii/S2589004218300816 (visited on 11/16/2018).

Chen, Xingqi et al. (2016). "ATAC-See Reveals the Accessible Genome by Transposase-Mediated Imaging and Sequencing". In: *Nature Methods* 13.12. 00000, pp. 1013–1020. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4031. URL: http://www.nature.com/doifinder/10.1038/nmeth.4031 (visited on 02/06/2017).

Chen, Yinghua et al. (2011). "C-Myc Activates BRCA1 Gene Expression through Distal Promoter Elements in Breast Cancer Cells". In: *BMC cancer* 11.1. 00017, p. 1. URL: http://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-11-246 (visited on 08/23/2016).

Chen, Yu et al. (2018). "Mapping 3D Genome Organization Relative to Nuclear Compartments Using TSA-Seq as a Cytological Ruler". In: *J Cell Biol* 217.11, pp. 4025–4048. ISSN: 0021-9525, 1540-8140. DOI: 10.1083/jcb.201807108. pmid: 30154186. URL: http://jcb.rupress.org/content/217/11/4025 (visited on 11/16/2018).

Chenouard, Nicolas et al. (2014). "Objective Comparison of Particle Tracking Methods". In: *Nature Methods* 11.3. 00211, pp. 281–289. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.2808. URL: http://www.nature.com/doifinder/10.1038/nmeth.2808 (visited on 01/05/2017).

Chereji, Răzvan V. and David J. Clark (2018). "Major Determinants of Nucleosome Positioning". In: *Biophysical Journal* 114.10, pp. 2279–2289. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2018.03.015. pmid: 29628211. URL: https://www.cell.com/biophysj/abstract/S0006-3495(18)30381-3 (visited on 04/06/2019).

Cho, Won-Ki, Namrata Jayanth, Brian P. English, et al. (2016). "RNA Polymerase II Cluster Dynamics Predict mRNA Output in Living Cells". In: *eLife* 5, e13617. ISSN: 2050-084X. DOI: 10.7554/eLife.13617. URL: https://elifesciences.org/articles/13617 (visited on 07/13/2018).

Cho, Won-Ki, Namrata Jayanth, Susan Mullen, et al. (2016). "Super-Resolution Imaging of Fluorescently Labeled, Endogenous RNA Polymerase II in Living Cells with CRISPR/Cas9-Mediated Gene Editing". In: *Scientific Reports* 6.1. 00000, p. 35949. ISSN: 2045-2322. DOI: 10.1038/srep35949. URL: http://www.nature.com/articles/srep35949 (visited on 06/18/2017).

Chong, Shasha, Claire Dugast-Darzacq, Zhe Liu, Peng Dong, Gina M. Dailey, et al. (2018). "Imaging Dynamic and Selective Low-Complexity Domain Interactions That Control Gene Transcription". In: *Science*, eaar2555. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aar2555. pmid: 29930090. URL: http://science.sciencemag.org/content/early/2018/06/20/science.aar2555 (visited on 06/25/2018).

Chong, Shasha, Claire Dugast-Darzacq, Zhe Liu, Peng Dong, Gina Dailey, et al. (2017). "Dynamic and Selective Low-Complexity Domain Interactions Revealed by Live-Cell Single-Molecule Imaging". In: *bioRxiv*, p. 208710. DOI: 10.1101/208710. URL: https://www.biorxiv.org/content/early/2017/10/25/208710 (visited on 10/26/2017).

Chronis, Constantinos et al. (2017). "Cooperative Binding of Transcription Factors Orchestrates Reprogramming". In: *Cell* 168.3. 00010, 442–459.e20. ISSN: 00928674. DOI: 10.1016/j.cell.2016.12.016. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867416317305 (visited on 07/17/2017).

Cisse, I. I. et al. (2013). "Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells". In: *Science* 341.6146, pp. 664–667. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1239053. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1239053 (visited on 11/21/2017).

Clauß, Karen et al. (2017). "DNA Residence Time Is a Regulatory Factor of Transcription Repression". In: *Nucleic Acids Research*. DOI: 10.1093/nar/gkx728. URL: https://academic.oup.com/nar/article/doi/10.1093/nar/gkx728/4085842 (visited on 10/27/2017).

Clavería, Cristina et al. (2013). "Myc-Driven Endogenous Cell Competition in the Early Mammalian Embryo". In: *Nature* 500.7460, pp. 39–44. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature12389. URL: http://www.nature.com/doifinder/10.1038/nature12389 (visited on 11/18/2017).

Coeurjolly, Jean-François (2007). *Simulation and Identification of the Fractional Brownian Motion: A Bibliographical and Comparative Study*. P.

Cohen, A. E. and W. E. Moerner (2006). "Suppressing Brownian Motion of Individual Biomolecules in Solution". In: *Proceedings of the National Academy of Sciences* 103.12, pp. 4362–4365. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0509976103. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0509976103 (visited on 04/25/2019).

Cohen, Adam E. and W. E. Moerner (2005). "Method for Trapping and Manipulating Nanoscale Objects in Solution". In: *Applied Physics Letters* 86.9, p. 093109. ISSN: 0003-6951, 1077-3118. DOI: 10.1063/1.1872220. URL: http://aip.scitation.org/doi/10.1063/1.1872220 (visited on 04/25/2019).

Cohen, Edward A. K., Anish V. Abraham, and Raimund J. Ober (2017). "Resolution Limit of Image Analysis Algorithms". In: *bioRxiv*, p. 240531. DOI: 10.1101/240531. URL: https://www.biorxiv.org/content/early/2017/12/28/240531 (visited on 12/29/2017).

Coleman, Robert A. et al. (2017). "P53 Dynamically Directs TFIID Assembly on Target Gene Promoters". In: *bioRxiv*, p. 083014. DOI: 10.1101/083014. URL: https://www.biorxiv.org/content/early/2017/02/24/083014 (visited on 09/20/2017).

Condamin, S., O. Bénichou, et al. (2007). "First-Passage Times in Complex Scale-Invariant Media". In: *Nature* 450.7166, pp. 77–80. ISSN: 0028-0836. DOI: 10.1038/nature06201. URL: http://www.nature.com/nature/journal/v450/n7166/full/nature06201.html (visited on 10/29/2017).

Condamin, S., Vincent Tejedor, et al. (2008). "Probing Microscopic Origins of Confined Subdiffusion by First-Passage Observables". In: *Proceedings of the National Academy of Sciences* 105.15, pp. 5675–5680.

Coppey, M. et al. (2004). "Kinetics of Target Site Localization of a Protein on DNA: A Stochastic Approach., Kinetics of Target Site Localization of a Protein on DNA: A Stochastic Approach". In: *Biophysical journal, Biophysical Journal* 87, 87.3, 3, pp. 1640, 1640–1649. ISSN: 0006-3495. DOI: 10.1529/biophysj.104.045773,10.1529/biophysj.104.045773. pmid: 15345543. URL: http://europepmc.org/abstract/MED/15345543,%20http://europepmc.org/articles/PMC1304569/?report=abstract (visited on 10/29/2017).

Cortini, Ruggero and Guillaume J. Filion (2018). "Theoretical Principles of Transcription Factor Traffic on Folded Chromatin". In: *Nature Communications* 9.1. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04130-x. URL: http://www.nature.com/articles/s41467-018-04130-x (visited on 06/12/2018).

Cotterman, R. et al. (2008). "N-Myc Regulates a Widespread Euchromatic Program in the Human Genome Partially Independent of Its Role as a Classical Transcription Factor". In: *Cancer Research* 68.23. 00110, pp. 9654–9662. ISSN: 0008-5472, 1538-7445. DOI: 10.1158/0008-5472.CAN-08-1961. URL: http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-08-1961 (visited on 08/23/2016).

Cowling, V. H. et al. (2006). "A Conserved Myc Protein Domain, MBIV, Regulates DNA Binding, Apoptosis, Transformation, and G2 Arrest". In: *Molecular and Cellular Biology* 26.11, pp. 4226–4239. ISSN: 0270-7306. DOI: 10.1128/MCB.01959-05. URL: http://mcb.asm.org/cgi/doi/10.1128/MCB.01959-05 (visited on 08/23/2016).

Craig, Cecil C. (1936). "On the Frequency Function of xy". In: *The Annals of Mathematical Statistics* 7.1, pp. 1–15. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177732541. URL: https://projecteuclid.org/euclid.aoms/1177732541 (visited on 10/28/2017).

Cremer, Thomas and Marion Cremer (2010). "Chromosome Territories". In: *Cold Spring Harbor Perspectives in Biology* 2.3, a003889. ISSN: , 1943-0264. DOI: 10.1101/cshperspect.a003889. pmid: 20300217. URL: http://cshperspectives.cshlp.org/content/2/3/a003889 (visited on 06/12/2018).

Cremer, Thomas, Marion Cremer, et al. (2015). "The 4D Nucleome: Evidence for a Dynamic Nuclear Landscape Based on Co-Aligned Active and Inactive Nuclear Compartments". In: *FEBS Letters* 589 (20PartA). 00049, pp. 2931–2943. ISSN: 00145793. DOI: 10.1016/j.febslet.2015.05.037. URL: http://doi.wiley.com/10.1016/j.febslet.2015.05.037 (visited on 04/02/2017).

Crocker, Justin et al. (2015). "Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness". In: *Cell* 160.1, pp. 191–203. ISSN: 0092-8674. DOI: 10.1016/j.cell.2014.11.041. URL: http://www.sciencedirect.com/science/article/pii/S0092867414015189 (visited on 05/24/2018).

Croft, Jenny A. et al. (1999). "Differences in the Localization and Morphology of Chromosomes in the Human Nucleus". In: *The Journal of Cell Biology* 145.6, pp. 1119–1131. ISSN: 0021-9525. pmid: 10366586. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2133153/ (visited on 04/06/2019).

Culley, Siân et al. (2017). "NanoJ-SQUIRREL: Quantitative Mapping and Minimisation of Super-Resolution Optical Imaging Artefacts". In: *bioRxiv*. 00000. DOI: 10.1101/158279. URL: http://biorxiv.org/content/early/2017/07/02/158279.abstract.

Daddysman, Matthew K. and Christopher J. Fecko (2013). "Revisiting Point FRAP to Quantitatively Characterize Anomalous Diffusion in Live Cells". In: *The Journal of Physical Chemistry B* 117.5, pp. 1241–1251. ISSN: 1520-6106. DOI: 10.1021/jp310348s. URL: http://dx.doi.org/10.1021/jp310348s (visited on 09/12/2017).

Dagliyan, O. et al. (2013). "Rational Design of a Ligand-Controlled Protein Conformational Switch". In: *Proceedings of the National Academy of Sciences* 110.17. 00028, pp. 6800–6804. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1218319110. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1218319110 (visited on 09/29/2016).

Dalla-Favera, R. et al. (1982). "Human C-Myc Onc Gene Is Located on the Region of Chromosome 8 That Is Translocated in Burkitt Lymphoma Cells". In: *Proceedings of the National Academy of Sciences* 79.24, pp. 7824–7827. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.79.24.7824. pmid: 6961453. URL: https://www.pnas.org/content/79/24/7824 (visited on 03/30/2019).

Dang, Chi V. (1999). "C-Myc Target Genes Involved in Cell Growth, Apoptosis, and Metabolism". In: *Molecular and cellular biology* 19.1. 01298, pp. 1–11. URL: http://mcb.asm.org/content/19/1/1.short (visited on 08/25/2016).

— (2012). "MYC on the Path to Cancer". In: *Cell* 149.1, pp. 22–35. ISSN: 00928674. DOI: 10.1016/j.cell.2012.03.003. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867412002966 (visited on 11/13/2018).

Dang, Chi V. et al. (2006). "The C-Myc Target Gene Network". In: *Seminars in Cancer Biology* 16.4. 00644, pp. 253–264. ISSN: 1044579X. DOI: 10.1016/j.semcancer.2006.07.014. URL: http://linkinghub.elsevier.com/retrieve/pii/S1044579X06000654 (visited on 08/23/2016).

Darzacq, Xavier et al. (2007). "In Vivo Dynamics of RNA Polymerase II Transcription". In: *Nature Structural & Molecular Biology* 14.9, pp. 796–806. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/nsmb1280. URL: http://www.nature.com/doifinder/10.1038/nsmb1280 (visited on 09/29/2017).

Das, Rahul K., Scott L. Crick, and Rohit V. Pappu (2012). "N-Terminal Segments Modulate the α-Helical Propensities of the Intrinsically Disordered Basic Regions of bZIP Proteins". In: *Journal of Molecular Biology* 416.2, pp. 287–299. ISSN: 00222836. DOI: 10.1016/j.jmb.2011.12.043. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022283611013726 (visited on 11/05/2017).

Das, Rahul K, Kiersten M Ruff, and Rohit V Pappu (2015). "Relating Sequence Encoded Information to Form and Function of Intrinsically Disordered Proteins". In: *Current Opinion in Structural Biology* 32, pp. 102–112. ISSN: 0959440X. DOI: 10.1016/j.sbi.2015.03.008. URL: http://linkinghub.elsevier.com/retrieve/pii/S0959440X15000354 (visited on 11/05/2017).

Davidson, Iain F. et al. (2016). "Rapid Movement and Transcriptional Re-localization of Human Cohesin on DNA". In: *The EMBO Journal* 35.24, pp. 2671–2685. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.201695402. pmid: 27799150. URL: http://emboj.embopress.org/content/35/24/2671 (visited on 04/06/2019).

De Gennes, P. G. (1982). "Kinetics of Diffusion-controlled Processes in Dense Polymer Systems. I. Nonentangled Regimes". In: *The Journal of Chemical Physics* 76.6, pp. 3316–3321. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.443328. URL: http://aip.scitation.org/doi/10.1063/1.443328 (visited on 11/22/2017).

De Pretis, Stefano et al. (2017). "Integrative Analysis of RNA Polymerase II and Transcriptional Dynamics upon MYC Activation". In: *Genome Research*. ISSN: 1549-5469. DOI: 10.1101/gr.226035.117. pmid: 28904013.

Dejure, Francesca R and Martin Eilers (2017). "MYC and Tumor Metabolism: Chicken and Egg". In: *The EMBO Journal*, e201796438. ISSN: 0261-4189, 1460-2075, 0261-4189, 1460-2075. DOI: 10.15252/embj.201796438. URL: http://emboj.embopress.org/lookup/doi/10.15252/embj.201796438 (visited on 11/12/2017).

Dekker, Job et al. (2002). "Capturing Chromosome Conformation". In: *Science* 295.5558, pp. 1306–1311. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1067799. pmid: 11847345. URL: http://science.sciencemag.org/content/295/5558/1306 (visited on 04/06/2019).

Delachat, Aurore M.-F. et al. (2018). "Engineered Multivalent Sensors to Detect Coexisting Histone Modifications in Living Stem Cells". In: *Cell Chemical Biology* 25.1, 51–56.e6. ISSN: 24519456. DOI: 10.1016/j.chembiol.2017.10.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S2451945617303896 (visited on 10/16/2018).

Deng, Wulan et al. (2014). "Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping". In: *Cell* 158.4, pp. 849–860. ISSN: 0092-8674. DOI: 10.1016/j.cell.2014.05.050. URL: http://www.sciencedirect.com/science/article/pii/S0092867414008113 (visited on 01/16/2018).

Deschout, Hendrik, Kristiaan Neyts, and Kevin Braeckmans (2012). "The Influence of Movement on the Localization Precision of Sub-Resolution Particles in Fluorescence Microscopy". In: *Journal of Biophotonics* 5.1, pp. 97–109. ISSN: 1864063X. DOI: 10.1002/jbio.201100078. URL: http://doi.wiley.com/10.1002/jbio.201100078 (visited on 10/14/2017).

Díaz-Díaz, Covadonga et al. (2017). "Pluripotency Surveillance by Myc-Driven Competitive Elimination of Differentiating Cells". In: *Developmental Cell*. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2017.08.011. URL: http://www.cell.com/developmental-cell/abstract/S1534-5807(17)30671-8 (visited on 09/17/2017).

Dieker, Ton (2004). "Simulation of Fractional Brownian Motion". In: *MSc theses, University of Twente, Amsterdam, The Netherlands*.

Dingar, Dharmendra et al. (2015). "BioID Identifies Novel C-MYC Interacting Partners in Cultured Cells and Xenograft Tumors". In: *Journal of Proteomics* 118. 00028, pp. 95–111. ISSN: 18743919. DOI: 10.1016/j.jprot.2014.09.029. URL: http://linkinghub.elsevier.com/retrieve/pii/S187439191400462X (visited on 04/28/2017).

Dixon, Jesse R. et al. (2012). "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions". In: *Nature* 485.7398. 01371, pp. 376–380. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature11082. URL: http://www.nature.com/doifinder/10.1038/nature11082 (visited on 01/11/2017).

Dogan, Nergiz et al. (2015). "Occupancy by Key Transcription Factors Is a More Accurate Predictor of Enhancer Activity than Histone Modifications or Chromatin Accessibility". In: *Epigenetics & Chromatin* 8, p. 16. ISSN: 1756-8935. DOI: 10.1186/s13072-015-0009-5. URL: https://doi.org/10.1186/s13072-015-0009-5 (visited on 01/25/2018).

Doi, Masao and Samuel F. Edwards (2007). *The Theory of Polymer Dynamics*. Reprinted. International Series of Monographs on Physics 73. OCLC: 845169495. Oxford: Clarendon Press. 391 pp. ISBN: 978-0-19-852033-7.

Donaghey, Julie et al. (2018). "Genetic Determinants and Epigenetic Effects of Pioneer-Factor Occupancy". In: *Nature Genetics* 50.2, pp. 250–258. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/s41588-017-0034-3. URL: http://www.nature.com/articles/s41588-017-0034-3 (visited on 06/14/2018).

Dorigo, Benedetta et al. (2004). "Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber". In: *Science* 306.5701, pp. 1571–1573. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1103124. pmid: 15567867. URL: http://science.sciencemag.org/content/306/5701/1571 (visited on 04/06/2019).

Doucleff, Michaeleen and G. Marius Clore (2008). "Global Jumping and Domain-Specific Intersegment Transfer between DNA Cognate Sites of the Multidomain Transcription Factor Oct-1". In: *Proceedings of the National Academy of Sciences* 105.37, pp. 13871–13876.

Douglass, Kyle M. (2017). *Implementing a Fast Gibson-Lanni PSF Solver in Python*. URL: http://kmdouglass.github.io/posts/implementing-a-fast-gibson-lanni-psf-solver-in-python.html (visited on 11/23/2017).

Dowen, Jill M. et al. (2013). "Multiple Structural Maintenance of Chromosome Complexes at Transcriptional Regulatory Elements". In: *Stem Cell Reports* 1.5, pp. 371–378. ISSN: 22136711. DOI: 10.1016/j.stemcr.2013.09.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S2213671113000891 (visited on 04/07/2019).

Drawert, Brian et al. (2016). "Stochastic Simulation Service: Bridging the Gap between the Computational Expert and the Biologist". In: *PLOS Computational Biology* 12.12, e1005220. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005220. URL: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005220 (visited on 10/20/2017).

Du, Zhenhai et al. (2017). "Allelic Reprogramming of 3D Chromatin Architecture during Early Mammalian Development". In: *Nature* 547.7662, pp. 232–235. ISSN: 1476-4687. DOI: 10.1038/nature23263. URL: https://www.nature.com/articles/nature23263 (visited on 06/11/2018).

Dunsing, Valentin et al. (2018). "Quantifying Protein Oligomerization in Living Cells: A Systematic Comparison of Fluorescent Proteins". In: *BioRxiv*. DOI: 10.1101/311175. URL: http://biorxiv.org/lookup/doi/10.1101/311175 (visited on 06/26/2018).

Durand, Neva C. et al. (2016). "Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom". In: *Cell Systems* 3.1, pp. 99–101. ISSN: 2405-4712. DOI: 10.1016/j.cels.2015.07.012. URL: http://www.sciencedirect.com/science/article/pii/S240547121500054X (visited on 04/06/2019).

Eberhardy, S. R., C. A. D'Cunha, and P. J. Farnham (2000). "Direct Examination of Histone Acetylation on Myc Target Genes Using Chromatin Immunoprecipitation". In: *Journal of Biological Chemistry* 275.43. 00099, pp. 33798–33805. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M005154200. URL: http://www.jbc.org/cgi/doi/10.1074/jbc.M005154200 (visited on 12/09/2016).

El Beheiry, Mohamed, Maxime Dahan, and Jean-Baptiste Masson (2015). "InferenceMAP: Mapping of Single-Molecule Dynamics with Bayesian Inference". In: *Nature Methods* 12.7. 00009, pp. 594–595. URL: http://www.nature.com/nmeth/journal/v12/n7/full/nmeth.3441.html (visited on 08/28/2016).

El Beheiry, Mohamed, Silvan Türkcan, et al. (2016). "A Primer on the Bayesian Approach to High-Density Single-Molecule Trajectories Analysis". In: *Biophysical Journal* 110.6. 00000, pp. 1209–1215. ISSN: 00063495. DOI: 10.1016/

j.bpj.2016.01.018. URL: http://linkinghub.elsevier.com/retrieve/pii/S000634951600093X (visited on 05/24/2016).

Elf, J., G.-W. Li, and X. S. Xie (2007). "Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell". In: *Science* 316.5828, pp. 1191–1194. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1141967. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1141967 (visited on 11/21/2017).

Enderlein, Jörg, Erdal Toprak, and Paul R. Selvin (2006). "Polarization Effect on Position Accuracy of Fluorophore Localization". In: *Optics Express* 14.18, pp. 8111–8120. ISSN: 1094-4087. DOI: 10.1364/OE.14.008111. URL: https://www.osapublishing.org/abstract.cfm?uri=oe-14-18-8111 (visited on 12/05/2017).

Eng, Chee-Huat Linus et al. (2019). "Transcriptome-Scale Super-Resolved Imaging in Tissues by RNA seqFISH+". In: *Nature*, p. 1. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1049-y. URL: https://www.nature.com/articles/s41586-019-1049-y (visited on 03/30/2019).

Erdel, Fabian and Karsten Rippe (2018). "Formation of Chromatin Subcompartments by Phase Separation". In: *Biophysical Journal* 114.10, pp. 2262–2270. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2018.03.011. pmid: 29628210. URL: https://www.cell.com/biophysj/abstract/S0006-3495(18)30349-7 (visited on 06/27/2018).

Ernst, Dominique, Jürgen Köhler, and Matthias Weiss (2014). "Probing the Type of Anomalous Diffusion with Single-Particle Tracking". In: *Phys. Chem. Chem. Phys.* 16.17. 00038, pp. 7686–7691. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/C4CP00292J. URL: http://xlink.rsc.org/?DOI=C4CP00292J (visited on 06/18/2017).

Etheridge, Thomas J. et al. (2014). "Quantification of DNA-Associated Proteins inside Eukaryotic Cells Using Single-Molecule Localization Microscopy". In: *Nucleic Acids Research* 42.19, e146–e146. ISSN: 0305-1048. DOI: 10.1093/nar/gku726. URL: https://academic.oup.com/nar/article/42/19/e146/2902489/Quantification-of-DNA-associated-proteins-inside (visited on 10/20/2017).

Etoc, Fred et al. (2018). "Non-Specific Interactions Govern Cytosolic Diffusion of Nanosized Objects in Mammalian Cells". In: *Nature Materials* 17.8, pp. 740–746. ISSN: 1476-4660. DOI: 10.1038/s41563-018-0120-7. URL: https://www.nature.com/articles/s41563-018-0120-7 (visited on 09/22/2018).

Fagnocchi, Luca and Alessio Zippo (2017). "Multiple Roles of MYC in Integrating Regulatory Networks of Pluripotent Stem Cells". In: *Frontiers in Cell and Developmental Biology* 5. ISSN: 2296-634X. DOI: 10.3389/fcell.2017.00007. URL: https://www.frontiersin.org/articles/10.3389/fcell.2017.00007/full (visited on 08/28/2018).

Falk, Martin et al. (2018). "Heterochromatin Drives Organization of Conventional and Inverted Nuclei". In: *bioRxiv*, p. 244038. DOI: 10.1101/244038. URL: https://www.biorxiv.org/content/early/2018/01/09/244038 (visited on 01/31/2018).

Fan-Minogue, Hua et al. (2010). "Noninvasive Molecular Imaging of C-Myc Activation in Living Mice". In: *Proceedings of the National Academy of Sciences* 107.36, pp. 15892–15897. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1007443107. pmid: 20713710. URL: http://www.pnas.org/content/107/36/15892 (visited on 10/07/2017).

Fast and Accurate Three-Dimensional Point Spread Function Computation for Fluorescence Microscopy - Jizhou Li's Homepage (2017). URL: http://www.ee.cuhk.edu.hk/~jzli/MicroscPSF/ (visited on 11/23/2017).

Fazel, Mohamadreza et al. (2019). "Bayesian Multiple Emitter Fitting Using Reversible Jump Markov Chain Monte Carlo". In: *bioRxiv*, p. 530261. DOI: 10.1101/530261. URL: https://www.biorxiv.org/content/10.1101/530261v1 (visited on 03/05/2019).

Fei, Jingyi et al. (2017). "Quantitative Analysis of Multilayer Organization of Proteins and RNA in Nuclear Speckles at Super Resolution". In: *J Cell Sci*, jcs–206854.

Feng, Mudong and Michael K. Gilson (2019). "A Thermodynamic Limit on the Role of Self-Propulsion in Enhanced Enzyme Diffusion". In: *Biophysical Journal*, S0006349519303054. ISSN: 00063495. DOI: 10.1016/j.bpj.2019.04.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0006349519303054 (visited on 04/25/2019).

Feric, Marina et al. (2016). "Coexisting Liquid Phases Underlie Nucleolar Subcompartments". In: *Cell* 165.7, pp. 1686–1697. ISSN: 00928674. DOI: 10.1016/j.cell.2016.04.047. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867416304925 (visited on 05/24/2018).

Fernandez, P. C. et al. (2003). "Genomic Targets of the Human C-Myc Protein". In: *Genes & Development* 17.9. 00780, pp. 1115–1129. ISSN: 08909369. DOI: 10.1101/gad.1067003. URL: http://www.genesdev.org/cgi/doi/10.1101/gad.1067003 (visited on 08/23/2016).

Festuccia, Nicola, Agnès Dubois, et al. (2016). "Mitotic Binding of Esrrb Marks Key Regulatory Regions of the Pluripotency Network". In: *Nature Cell Biology* 18.11, p. 1139. ISSN: 1476-4679. DOI: 10.1038/ncb3418. URL: https://www.nature.com/articles/ncb3418 (visited on 01/26/2018).

Festuccia, Nicola, Nick Owens, et al. (2018). "Transcription Factor Activity and Nucleosome Organisation in Mitosis". In: *bioRxiv*, p. 392241. DOI: 10.1101/392241. URL: https://www.biorxiv.org/content/early/2018/08/15/392241 (visited on 08/16/2018).

Fieber, Wolfgang et al. (2001). "Structure, Function, and Dynamics of the Dimerization and DNA-Binding Domain of Oncogenic Transcription Factor v-Myc". In: *Journal of Molecular Biology* 307.5, pp. 1395–1410. ISSN: 00222836. DOI: 10.1006/jmbi.2001.4537. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022283601945371 (visited on 07/05/2018).

Fields, Brandon D et al. (2018). "A Multiplexed DNA FISH Strategy for Assessing Genome Architecture in C. Elegans". In: DOI: 10.1101/397471. URL: http://biorxiv.org/lookup/doi/10.1101/397471 (visited on 09/27/2018).

Filipczyk, Adam et al. (2015). "Network Plasticity of Pluripotency Transcription Factors in Embryonic Stem Cells". In: *Nature Cell Biology* 17.10, ncb3237. ISSN: 1476-4679. DOI: 10.1038/ncb3237. URL: https://www.nature.com/articles/ncb3237 (visited on 10/26/2017).

Finch, J T and A Klug (1976). "Solenoidal Model for Superstructure in Chromatin." In: *Proceedings of the National Academy of Sciences of the United States of America* 73.6, pp. 1897–1901. ISSN: 0027-8424. pmid: 1064861. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC430414/ (visited on 04/06/2019).

Fitz, Veronika et al. (2016). "Nucleosomal Arrangement Affects Single-Molecule Transcription Dynamics". In: *Proceedings of the National Academy of Sciences* 113.45, pp. 12733–12738. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1602764113. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1602764113 (visited on 04/10/2019).

Flavahan, William A. et al. (2016). "Insulator Dysfunction and Oncogene Activation in *IDH* Mutant Gliomas". In: *Nature* 529.7584, pp. 110–114. ISSN: 1476-4687. DOI: 10.1038/nature16490. URL: https://www.nature.com/articles/nature16490 (visited on 04/06/2019).

Fletcher, Steven and Edward V. Prochownik (2015). "Small-Molecule Inhibitors of the Myc Oncoprotein". In: *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. Myc Proteins in Cell Biology and Pathology 1849.5, pp. 525–543. ISSN: 1874-9399. DOI: 10.1016/j.bbagrm.2014.03.005. URL: http://www.sciencedirect.com/science/article/pii/S1874939914000571 (visited on 11/13/2018).

Florens-Zmirou, Danielle (1993). "On Estimating the Diffusion Coefficient from Discrete Observations". In: *Journal of Applied Probability* 30.4. 00274, p. 790. ISSN: 00219002. DOI: 10.2307/3214513. JSTOR: 3214513?origin=crossref.

Flyamer, Ilya M. et al. (2017). "Single-Nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-to-Zygote Transition". In: *Nature* 544.7648, pp. 110–114. ISSN: 1476-4687. DOI: 10.1038/nature21711. URL: https://www.nature.com/articles/nature21711 (visited on 04/06/2019).

Frank, Scott R. et al. (2001). "Binding of C-Myc to Chromatin Mediates Mitogen-Induced Acetylation of Histone H4 and Gene Activation". In: *Genes & development* 15.16. 00446, pp. 2069–2082. URL: http://genesdev.cshlp.org/content/15/16/2069.short (visited on 08/23/2016).

Freedberg, Darón I. and Philipp Selenko (2014). "Live Cell NMR". In: *Annual Review of Biophysics* 43.1, pp. 171–192. ISSN: 1936-122X, 1936-1238. DOI: 10.1146/annurev-biophys-051013-023136. URL: http://www.annualreviews.org/doi/10.1146/annurev-biophys-051013-023136 (visited on 10/30/2017).

Frost, Nicholas A., Hsiangmin E. Lu, and Thomas A. Blanpied (2012). "Optimization of Cell Morphology Measurement via Single-Molecule Tracking PALM". In: *PloS One* 7.5, e36751. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0036751. pmid: 22570741.

Fudenberg, Geoffrey, Nezar Abdennur, et al. (2018). "Emerging Evidence of Chromosome Folding by Loop Extrusion". In: *bioRxiv*, p. 264648. DOI: 10.1101/264648. URL: https://www.biorxiv.org/content/early/2018/02/16/264648 (visited on 02/17/2018).

Fudenberg, Geoffrey, Maxim Imakaev, et al. (2016). "Formation of Chromosomal Domains by Loop Extrusion". In: *Cell Reports* 15.9. 00068, pp. 2038–2049. ISSN: 22111247. DOI: 10.1016/j.celrep.2016.04.085. URL: http://linkinghub.elsevier.com/retrieve/pii/S2211124716305307 (visited on 04/09/2017).

Fudenberg, Geoffrey and Leonid A Mirny (2012). "Higher-Order Chromatin Structure: Bridging Physics and Biology". In: *Current Opinion in Genetics & Development*. Genome Architecture and Expression 22.2, pp. 115–124. ISSN: 0959-437X. DOI: 10.1016/j.gde.2012.01.006. URL: http://www.sciencedirect.com/science/article/pii/S0959437X1200007X (visited on 04/06/2019).

Fulco, Charles P. et al. (2016). "Systematic Mapping of Functional Enhancer–Promoter Connections with CRISPR Interference". In: *Science* 354.6313, pp. 769–773. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aag2445. pmid: 27708057. URL: http://science.sciencemag.org/content/354/6313/769 (visited on 03/06/2019).

Fussner, Eden et al. (2011). "Constitutive Heterochromatin Reorganization during Somatic Cell Reprogramming". In: *The EMBO Journal* 30.9, pp. 1778–1789. ISSN: 0261-4189, 1460-2075. DOI: 10.1038/emboj.2011.96. pmid: 21468033. URL: http://emboj.embopress.org/content/30/9/1778 (visited on 04/06/2019).

Gaj, Thomas, Charles A. Gersbach, and Carlos F. Barbas (2013). "ZFN, TALEN, and CRISPR/Cas-Based Methods for Genome Engineering". In: *Trends in biotechnology* 31.7. 01134, pp. 397–405. URL: http://www.sciencedirect.com/science/article/pii/S0167779913000875 (visited on 08/23/2016).

Gambarotto, Davide et al. (2018). "Imaging beyond the Super-Resolution Limits Using Ultrastructure Expansion Microscopy (UltraExM)". In: DOI: 10.1101/308270. URL: http://biorxiv.org/lookup/doi/10.1101/308270 (visited on 06/13/2018).

Ganai, Nirmalendu, Surajit Sengupta, and Gautam I. Menon (2014). "Chromosome Positioning from Activity-Based Segregation". In: *Nucleic Acids Research* 42.7, pp. 4145–4159. ISSN: 0305-1048. DOI: 10.1093/nar/gkt1417. URL: https://academic.oup.com/nar/article/42/7/4145/2436380 (visited on 04/06/2019).

Ganji, Mahipal et al. (2018). "Real-Time Imaging of DNA Loop Extrusion by Condensin". In: *Science*, eaar7831. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aar7831. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aar7831 (visited on 02/27/2018).

Gao, Mengfei et al. (2018). "Expansion Stimulated Emission Depletion Microscopy (ExSTED)". In: *bioRxiv*, p. 278937. DOI: 10.1101/278937. URL: https://www.biorxiv.org/content/early/2018/03/08/278937 (visited on 03/09/2018).

Gao, Tianshun et al. (2016). "EnhancerAtlas: A Resource for Enhancer Annotation and Analysis in 105 Human Cell/Tissue Types". In: *Bioinformatics*. 00000, btw495. ISSN: 1367-4803, 1460-2059. DOI: 10.1093/bioinformatics/

btw495. URL: http://bioinformatics.oxfordjournals.org/lookup/doi/10.1093/bioinformatics/btw495 (visited on 08/27/2016).

Gao, Yunfeng et al. (2017). "Charged Residues in the H-NS Linker Drive DNA Binding and Gene Silencing in Single Cells". In: *Proceedings of the National Academy of Sciences* 114.47, pp. 12560–12565. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1716721114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1716721114 (visited on 11/21/2017).

Garcia, Hernan G. et al. (2013). "Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning". In: *Current Biology* 23.21. 00044, pp. 2140–2145. ISSN: 09609822. DOI: 10.1016/j.cub.2013.08.054. URL: http://linkinghub.elsevier.com/retrieve/pii/S0960982213011135 (visited on 08/23/2016).

Garrison, Brian S. and Derrick J. Rossi (2010). "Controlling Stem Cell Fate One Substrate at a Time". In: *Nature immunology* 11.3. 00003, p. 193. URL: http://idi.harvard.edu/uploads/investigators/Garrison_Nat_Immunology_2010_copy.pdf (visited on 08/25/2016).

Gassler, Johanna et al. (2017). "A Mechanism of Cohesin-dependent Loop Extrusion Organizes Zygotic Genome Architecture". In: *The EMBO Journal* 36.24, pp. 3600–3618. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.201798083. pmid: 29217590. URL: http://emboj.embopress.org/content/36/24/3600 (visited on 04/06/2019).

Gebhardt, J. Christof M. et al. (2013). "Single-Molecule Imaging of Transcription Factor Binding to DNA in Live Mammalian Cells". In: *Nature Methods* 10.5, pp. 421–426. ISSN: 1548-7091. DOI: 10.1038/nmeth.2411. URL: https://www.nature.com/nmeth/journal/v10/n5/full/nmeth.2411.html (visited on 09/12/2017).

Gemayel, Rita et al. (2015). "Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity". In: *Molecular Cell* 59.4, pp. 615–627. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2015.07.003. URL: http://www.sciencedirect.com/science/article/pii/S1097276515005341 (visited on 02/01/2018).

Gerchman, S. E. and V. Ramakrishnan (1987). "Chromatin Higher-Order Structure Studied by Neutron Scattering and Scanning Transmission Electron Microscopy." In: *Proceedings of the National Academy of Sciences of the United States of America* 84.22, pp. 7802–7806. ISSN: 0027-8424. DOI: 10.1073/pnas.84.22.7802. pmid: 3479765. URL: https://europepmc.org/articles/PMC299397/ (visited on 04/06/2019).

Germier, Thomas et al. (2017). "Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement". In: *Biophysical Journal* 113.7, pp. 1383–1394. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2017.08.014. URL: http://www.sciencedirect.com/science/article/pii/S0006349517308731 (visited on 12/10/2018).

Ghosh, Rajarshi P et al. (2017). "A Fluorogenic Array Tag for Temporally Unlimited Single Molecule Tracking". In: *bioRxiv*. 00000. DOI: 10.1101/159004. URL: http://biorxiv.org/content/early/2017/07/03/159004.abstract.

Ghosh, Surya K., Andrey G. Cherstvy, and Ralf Metzler (2015). "Non-Universal Tracer Diffusion in Crowded Media of Non-Inert Obstacles". In: *Physical Chemistry Chemical Physics* 17.3, pp. 1847–1858. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/C4CP03599B. URL: http://xlink.rsc.org/?DOI=C4CP03599B (visited on 03/26/2018).

Gibbs, Eric B. and Scott A. Showalter (2016). "Quantification of Compactness and Local Order in the Ensemble of the Intrinsically Disordered Protein FCP1". In: *The Journal of Physical Chemistry B* 120.34, pp. 8960–8969. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/acs.jpcb.6b06934. URL: http://pubs.acs.org/doi/abs/10.1021/acs.jpcb.6b06934 (visited on 11/05/2017).

Gibcus, Johan H. et al. (2018). "A Pathway for Mitotic Chromosome Formation". In: *Science*, eaao6135. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aao6135. pmid: 29348367. URL: http://science.sciencemag.org/content/early/2018/01/17/science.aao6135 (visited on 03/16/2018).

Gibson, Daniel G et al. (2009). "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases". In: *Nature Methods* 6.5. 00000, pp. 343–345. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1318. URL: http://www.nature.com/doifinder/10.1038/nmeth.1318 (visited on 08/23/2016).

Gimona, Mario (2006). "Protein Linguistics — a Grammar for Modular Protein Assembly?" In: *Nature Reviews Molecular Cell Biology* 7.1, pp. 68–73. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/nrm1785. URL: http://www.nature.com/doifinder/10.1038/nrm1785 (visited on 11/21/2017).

Giuntoli, Rebecca D. et al. (2015). "DNA-Segment-Facilitated Dissociation of Fis and NHP6A from DNA Detected via Single-Molecule Mechanical Response". In: *Journal of Molecular Biology* 427.19, pp. 3123–3136. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2015.07.015. URL: http://www.sciencedirect.com/science/article/pii/S0022283615004167 (visited on 05/03/2018).

Golestanian, Ramin (2015). "Enhanced Diffusion of Enzymes That Catalyze Exothermic Reactions". In: *Physical Review Letters* 115.10, p. 108102. DOI: 10.1103/PhysRevLett.115.108102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.108102 (visited on 10/29/2017).

Goodman, Joseph W. (2005). *Introduction to Fourier Optics*. 3rd ed. OCLC: ocm56632414. Englewood, Colo: Roberts & Co. 491 pp. ISBN: 978-0-9747077-2-3.

Gorman, Jason et al. (2012). "Single-Molecule Imaging Reveals Target-Search Mechanisms during DNA Mismatch Repair". In: *Proceedings of the National Academy of Sciences* 109.45, E3074–E3083. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1211364109. pmid: 23012240. URL: http://www.pnas.org/content/109/45/E3074 (visited on 10/29/2017).

Goulian, Mark and Sanford M. Simon (2000). "Tracking Single Proteins within Cells". In: *Biophysical journal* 79.4, pp. 2188–2198.

Grauffel, Cédric, Roland H. Stote, and Annick Dejaegere (2010). "Force Field Parameters for the Simulation of Modified Histone Tails". In: *Journal of Computational Chemistry*, NA–NA. ISSN: 01928651, 1096987X. DOI: 10.1002/jcc.21536. URL: http://doi.wiley.com/10.1002/jcc.21536 (visited on 08/29/2018).

Grebenkov, Denis (2009). "Surveying Diffusion in Complex Geometries. An Essay". In: arXiv: 0909.1588 [math-ph]. URL: http://arxiv.org/abs/0909.1588 (visited on 04/16/2019).

Greenfeld, Max et al. (2015). "Single-Molecule Dataset (SMD): A Generalized Storage Format for Raw and Processed Single-Molecule Data". In: *BMC Bioinformatics* 16.1. ISSN: 1471-2105. DOI: 10.1186/s12859-014-0429-4. pmid: 25591752. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4384321/ (visited on 12/07/2018).

Greiss, Ferdinand et al. (2016). "Single-Molecule Imaging in Living Drosophila Embryos with Reflected Light-Sheet Microscopy". In: *Biophysical Journal* 110.4, pp. 939–946. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2015.12.035. URL: http://www.cell.com/biophysj/abstract/S0006-3495(16)00042-4 (visited on 10/30/2017).

Grimm, Jonathan B, Brian P English, Jiji Chen, et al. (2015). "A General Method to Improve Fluorophores for Live-Cell and Single-Molecule Microscopy". In: *Nature Methods* 12.3. 00066, pp. 244–250. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.3256. URL: http://www.nature.com/doifinder/10.1038/nmeth.3256 (visited on 07/28/2016).

Grimm, Jonathan B, Brian P English, Heejun Choi, et al. (2016). "Bright Photoactivatable Fluorophores for Single-Molecule Imaging". In: *Nature Methods* 13.12, pp. 985–988. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4034. URL: http://www.nature.com/articles/nmeth.4034 (visited on 04/25/2019).

Grimm, Jonathan B. et al. (2017). "A General Method to Fine-Tune Fluorophores for Live-Cell and in Vivo Imaging". In: *Nature Methods* 14.10, pp. 987–994. ISSN: 1548-7091. DOI: 10.1038/nmeth.4403. URL: http://www.nature.com/nmeth/journal/v14/n10/full/nmeth.4403.html (visited on 10/27/2017).

Grosberg, A. Yu. (2012). "How Two Meters of DNA Fit into a Cell Nucleus: Polymer Models with Topological Constraints and Experimental Data". In: *Polymer Science Series C* 54.1, pp. 1–10. ISSN: 1811-2382, 1555-614X. DOI: 10.1134/S1811238212070028. URL: http://link.springer.com/10.1134/S1811238212070028 (visited on 11/22/2017).

Grosberg, A. Yu, Sergei K. Nechaev, and Eugene I. Shakhnovich (1988). "The Role of Topological Constraints in the Kinetics of Collapse of Macromolecules". In: *Journal de physique* 49.12, pp. 2095–2100.

Grosberg, A. et al. (1993). "Crumpled Globule Model of the Three-Dimensional Structure of DNA". In: *EPL (Europhysics Letters)* 23.5, p. 373.

Grossman-Haham, Iris et al. (2018). "Slow Domain Reconfiguration Causes Power-Law Kinetics in a Two-State Enzyme". In: *Proceedings of the National Academy of Sciences* 115.3, pp. 513–518. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1714401115. pmid: 29298911. URL: http://www.pnas.org/content/115/3/513 (visited on 01/17/2018).

Grünwald, David, Andreas Hoekstra, et al. (2006). "Direct Observation of Single Protein Molecules in Aqueous Solution". In: *ChemPhysChem* 7.4, pp. 812–815. ISSN: 1439-7641. DOI: 10.1002/cphc.200500632. URL: http://onlinelibrary.wiley.com/doi/10.1002/cphc.200500632/abstract (visited on 10/03/2017).

Grünwald, David, Robert M. Martin, et al. (2008). "Probing Intranuclear Environments at the Single-Molecule Level". In: *Biophysical Journal* 94.7, pp. 2847–2858. ISSN: 0006-3495. DOI: 10.1529/biophysj.107.115014. pmid: 18065482. URL: http://www.cell.com/biophysj/abstract/S0006-3495(08)70535-6 (visited on 09/12/2017).

Guccione, Ernesto et al. (2006). "Myc-Binding-Site Recognition in the Human Genome Is Determined by Chromatin Context". In: *Nature Cell Biology* 8.7. 00000, pp. 764–770. ISSN: 1465-7392, 1476-4679. DOI: 10.1038/ncb1434. URL: http://www.nature.com/doifinder/10.1038/ncb1434 (visited on 08/23/2016).

Guelen, Lars et al. (2008). "Domain Organization of Human Chromosomes Revealed by Mapping of Nuclear Lamina Interactions". In: *Nature* 453.7197, pp. 948–951. ISSN: 1476-4687. DOI: 10.1038/nature06947. URL: https://www.nature.com/articles/nature06947 (visited on 04/06/2019).

Guigas, Gernot, Claudia Kalla, and Matthias Weiss (2007). "The Degree of Macromolecular Crowding in the Cytoplasm and Nucleoplasm of Mammalian Cells Is Conserved". In: *FEBS Letters* 581.26, pp. 5094–5098. ISSN: 00145793. DOI: 10.1016/j.febslet.2007.09.054. URL: http://doi.wiley.com/10.1016/j.febslet.2007.09.054 (visited on 09/22/2017).

Guigas, Gernot and Matthias Weiss (2008). "Sampling the Cell with Anomalous Diffusion—The Discovery of Slowness". In: *Biophysical Journal* 94.1, pp. 90–94. ISSN: 00063495. DOI: 10.1529/biophysj.107.117044. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349508707677 (visited on 11/19/2017).

Günther, Jan-Philipp, Michael Börsch, and Peer Fischer (2018). "Diffusion Measurements of Swimming Enzymes with Fluorescence Correlation Spectroscopy". In: *Accounts of Chemical Research* 51.9, pp. 1911–1920. ISSN: 0001-4842, 1520-4898. DOI: 10.1021/acs.accounts.8b00276. URL: http://pubs.acs.org/doi/10.1021/acs.accounts.8b00276 (visited on 04/25/2019).

Guo, Jiannan et al. (2014). "Sequence Specificity Incompletely Defines the Genome-Wide Occupancy of Myc". In: *Genome Biology*, p. 14.

Guo, Syuan-Ming et al. (2012). "Bayesian Approach to the Analysis of Fluorescence Correlation Spectroscopy Data II: Application to Simulated and In Vitro Data". In: *Analytical Chemistry* 84.9. 00033, pp. 3880–3888. ISSN: 0003-2700, 1520-6882. DOI: 10.1021/ac2034375. URL: http://pubs.acs.org/doi/abs/10.1021/ac2034375 (visited on 11/17/2016).

Gurskaya, Nadya G et al. (2006). "Engineering of a Monomeric Green-to-Red Photoactivatable Fluorescent Protein Induced by Blue Light". In: *Nature Biotechnology* 24.4. 00471, pp. 461–465. ISSN: 1087-0156. DOI: 10.1038/nbt1191. URL: http://www.nature.com/doifinder/10.1038/nbt1191 (visited on 08/23/2016).

Gustafson, William Clay et al. (2014). "Drugging MYCN through an Allosteric Transition in Aurora Kinase A". In: *Cancer Cell* 26.3, pp. 414–427. ISSN: 1535-6108. DOI: 10.1016/j.ccr.2014.07.015. URL: http://www.sciencedirect.com/science/article/pii/S1535610814003055 (visited on 05/28/2018).

Gustafsson, Nils et al. (2016). "Fast Live-Cell Conventional Fluorophore Nanoscopy with ImageJ through Super-Resolution Radial Fluctuations". In: *Nature Communications* 7. 00010, p. 12471. ISSN: 2041-1723. DOI: 10.1038/ncomms12471. URL: http://www.nature.com/doifinder/10.1038/ncomms12471 (visited on 07/20/2017).

Haarhuis, Judith H. I. et al. (2017). "The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension". In: *Cell* 169.4, 693–707.e14. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2017.04.013. pmid: 28475897. URL: https://www.cell.com/cell/abstract/S0092-8674(17)30426-9 (visited on 04/06/2019).

Haber, C., S. A. Ruiz, and D. Wirtz (2000). "Shape Anisotropy of a Single Random-Walk Polymer". In: *Proceedings of the National Academy of Sciences* 97.20, pp. 10792–10795. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.190320097. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.190320097 (visited on 04/17/2018).

Hadzhiev, Yavor et al. (2018). "A Cell Cycle-Coordinated Nuclear Compartment for Polymerase II Transcription Encompasses the Earliest Gene Expression before Global Genome Activation". In: DOI: 10.1101/366468. URL: http://biorxiv.org/lookup/doi/10.1101/366468 (visited on 10/16/2018).

Haile, Yohannes et al. (2014). "Characterization of the NT2-Derived Neuronal and Astrocytic Cell Lines as Alternative in Vitro Models for Primary Human Neurons and Astrocytes: NT2-Derived Neurons and Astrocytes". In: *Journal of Neuroscience Research* 92.9. 00006, pp. 1187–1198. ISSN: 03604012. DOI: 10.1002/jnr.23399. URL: http://doi.wiley.com/10.1002/jnr.23399 (visited on 08/23/2016).

Halazonetis, Thanos D. and Amany N. Kandil (1991). "Determination of the C-MYC DNA-Binding Site". In: *Proceedings of the National Academy of Sciences* 88.14. 00157, pp. 6162–6166. URL: http://www.pnas.org/content/88/14/6162.short (visited on 08/25/2016).

Hammar, Petter, Prune Leroy, et al. (2012). "The Lac Repressor Displays Facilitated Diffusion in Living Cells". In: *Science* 336.6088. 00093, pp. 1595–1598. ISSN: 0036-8075, 1095-9203.

Hammar, Petter, Mats Walldén, et al. (2014). "Direct Measurement of Transcription Factor Dissociation Excludes a Simple Operator Occupancy Model for Gene Regulation". In: *Nature Genetics* 46.4. 00046, pp. 405–408. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/ng.2905. URL: http://www.nature.com/doifinder/10.1038/ng.2905 (visited on 11/17/2016).

Han, Renmin et al. (2015). "Drift Correction for Single-Molecule Imaging by Molecular Constraint Field, a Distance Minimum Metric". In: *BMC Biophysics* 8.1. 00000, p. 1. ISSN: 2046-1682. DOI: 10.1186/s13628-014-0015-1. URL: http://www.biomedcentral.com/2046-1682/8/1 (visited on 07/20/2017).

Hann, Stephen R., Karen Sloan-Brown, and Gerald D. Spotts (1992). "Translational Activation of the Non-AUG-Initiated c-Myc 1 Protein at High Cell Densities Due to Methionine Deprivation." In: *Genes & development* 6.7. 00098, pp. 1229–1240. URL: http://genesdev.cshlp.org/content/6/7/1229.short (visited on 04/20/2017).

Hansen, Anders S., Assaf Amitai, et al. (2018). "Guided Nuclear Exploration Increases CTCF Target Search Efficiency". In: *bioRxiv*, p. 495457. DOI: 10.1101/495457. URL: https://www.biorxiv.org/content/10.1101/495457v1 (visited on 03/05/2019).

Hansen, Anders S., Iryna Pustova, et al. (2017). "CTCF and Cohesin Regulate Chromatin Loop Stability with Distinct Dynamics". In: *Elife* 6.

Hansen, Anders S., Maxime Woringer, et al. (2018). "Robust Model-Based Analysis of Single-Particle Tracking Experiments with Spot-On". In: *eLife* 7, e33125. ISSN: 2050-084X. DOI: 10.7554/eLife.33125. URL: https://elifesciences.org/articles/33125 (visited on 01/25/2018).

Hansen, Anders S et al. (2018). "An RNA-Binding Region Regulates CTCF Clustering and Chromatin Looping:" in: *bioRxiv*. DOI: 10.1101/495432. URL: http://biorxiv.org/lookup/doi/10.1101/495432 (visited on 02/27/2019).

Hansen, Jeffrey C. et al. (2018). "The 10-Nm Chromatin Fiber and Its Relationship to Interphase Chromosome Organization". In: *Biochemical Society Transactions* 46.1, pp. 67–76. ISSN: 1470-8752. DOI: 10.1042/BST20170101. pmid: 29263138.

Harmon, Tyler S. et al. (2016). "GADIS: Algorithm for Designing Sequences to Achieve Target Secondary Structure Profiles of Intrinsically Disordered Proteins". In: *Protein Engineering Design and Selection* 29.9, pp. 339–346. ISSN: 1741-0126, 1741-0134. DOI: 10.1093/protein/gzw034. URL: https://academic.oup.com/peds/article-lookup/doi/10.1093/protein/gzw034 (visited on 11/05/2017).

Hart, Jonathan R. et al. (2014). "Inhibitor of MYC Identified in a Kröhnke Pyridine Library". In: *Proceedings of the National Academy of Sciences* 111.34, pp. 12556–12561. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1319488111. pmid: 25114221. URL: http://www.pnas.org/content/111/34/12556 (visited on 07/03/2018).

Hartl, Markus (2016). "The Quest for Targets Executing MYC-Dependent Cell Transformation". In: *Frontiers in Oncology* 6. 00000. ISSN: 2234-943X. DOI: 10.3389/fonc.2016.00132. URL: http://journal.frontiersin.org/Article/10.3389/fonc.2016.00132/abstract (visited on 11/29/2016).

Hastreiter, Simon and Timm Schroeder (2016). "Nanog Dynamics in Single Embryonic Stem Cells". In: *Cell Cycle* 15.6, pp. 770–771. ISSN: 1538-4101. DOI: 10.1080/15384101.2015.1137711. pmid: 26760119. URL: http://dx.doi.org/10.1080/15384101.2015.1137711 (visited on 10/26/2017).

Havlin, Shlomo and Daniel Ben-Avraham (1987). "Diffusion in Disordered Media". In: *Advances in Physics* 36.6, pp. 695–798. ISSN: 0001-8732, 1460-6976. DOI: 10.1080/00018738700101072. URL: http://www.tandfonline.com/doi/abs/10.1080/00018738700101072 (visited on 08/23/2016).

He, Jun, Syuan-Ming Guo, and Mark Bathe (2012). "Bayesian Approach to the Analysis of Fluorescence Correlation Spectroscopy Data I: Theory". In: *Analytical Chemistry* 84.9. 00023, pp. 3871–3879. ISSN: 0003-2700, 1520-6882. DOI: 10.1021/ac2034369. URL: http://pubs.acs.org/doi/abs/10.1021/ac2034369 (visited on 11/17/2016).

He, Y. et al. (2008). "Random Time-Scale Invariant Diffusion and Transport Coefficients". In: *Physical Review Letters* 101.5. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.101.058101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.101.058101 (visited on 12/05/2017).

Hellenkamp, Björn et al. (2018). "Precision and Accuracy of Single-Molecule FRET Measurements—a Multi-Laboratory Benchmark Study". In: *Nature Methods* 15.9, p. 669. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0085-0. URL: https://www.nature.com/articles/s41592-018-0085-0 (visited on 11/27/2018).

Hellmann, Marcel et al. (2011). "Challenges in Determining Anomalous Diffusion in Crowded Fluids". In: *Journal of Physics: Condensed Matter* 23.23, p. 234113. ISSN: 0953-8984, 1361-648X. DOI: 10.1088/0953-8984/23/23/234113. URL: http://stacks.iop.org/0953-8984/23/i=23/a=234113?key=crossref.98fb782862bd1ed66893f26b1d52dbdf (visited on 12/05/2017).

Helmuth, Jo A., Grégory Paul, and Ivo F. Sbalzarini (2010). "Beyond Co-Localization: Inferring Spatial Interactions between Sub-Cellular Structures from Microscopy Images". In: *BMC Bioinformatics* 11, p. 372. ISSN: 1471-2105. DOI: 10.1186/1471-2105-11-372. URL: https://doi.org/10.1186/1471-2105-11-372 (visited on 11/18/2017).

Henriques, Ricardo et al. (2010). "QuickPALM: 3D Real-Time Photoactivation Nanoscopy Image Processing in ImageJ". In: *Nature Methods* 7.5, pp. 339–340. ISSN: 1548-7105. DOI: 10.1038/nmeth0510-339. URL: https://www.nature.com/articles/nmeth0510-339 (visited on 04/05/2019).

Herbert, Sébastien et al. (2017). "Chromatin Stiffening Underlies Enhanced Locus Mobility after DNA Damage in Budding Yeast". In: *The EMBO Journal*. 00000, e201695842. ISSN: 0261-4189, 1460-2075, 0261-4189, 1460-2075. DOI: 10.15252/embj.201695842. URL: http://emboj.embopress.org/lookup/doi/10.15252/embj.201695842 (visited on 07/19/2017).

Herbst, Andreas, Michael T Hemann, et al. (2005). "A Conserved Element in Myc That Negatively Regulates Its Proapoptotic Activity". In: *EMBO reports* 6.2. 00090, pp. 177–183. ISSN: 1469-221X, 1469-3178. DOI: 10.1038/sj.embor.7400333. URL: http://embor.embopress.org/cgi/doi/10.1038/sj.embor.7400333 (visited on 04/28/2017).

Herbst, Friederike, Claudia R Ball, et al. (2012). "Extensive Methylation of Promoter Sequences Silences Lentiviral Transgene Expression During Stem Cell Differentiation In Vivo". In: *Molecular Therapy* 20.5. 00029, pp. 1014–1021. ISSN: 1525-0016, 1525-0024. DOI: 10.1038/mt.2012.46. URL: http://www.nature.com/doifinder/10.1038/mt.2012.46 (visited on 09/29/2016).

Hernandez-Stumpfhauser, Daniel, F. Jay Breidt, and Mark J. van der Woerd (2017). "The General Projected Normal Distribution of Arbitrary Dimension: Modeling and Bayesian Inference". In: *Bayesian Analysis* 12.1, pp. 113–133. ISSN: 1936-0975. DOI: 10.1214/15-BA989. URL: http://projecteuclid.org/euclid.ba/1453211962 (visited on 10/29/2017).

Herrera, Francisco J et al. (2014). "Core Promoter Factor TAF9B Regulates Neuronal Gene Expression". In: *eLife* 3, e02559. ISSN: 2050-084X. DOI: 10.7554/eLife.02559. URL: https://elifesciences.org/articles/02559 (visited on 04/10/2019).

Hess, S. T., T. P. Girirajan, and M. D. Mason (2006). "Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy". In: *Biophys J* 91. 02222. DOI: 10.1529/biophysj.106.091116. URL: https://doi.org/10.1529/biophysj.106.091116.

Hettich, Johannes and J Christof M Gebhardt (2018). "Transcription Factor Target Site Search and Gene Regulation in a Background of Un- Specific Binding Sites". In: p. 25.

Heun, P. (2001). "Chromosome Dynamics in the Yeast Interphase Nucleus". In: *Science* 294.5549, pp. 2181–2186. ISSN: 00368075, 10959203. DOI: 10.1126/science.1065366. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1065366 (visited on 04/06/2019).

Hilbert, Lennart et al. (2017). "Transcription Establishes Microenvironments That Organize Euchromatin". In: *bioRxiv*, p. 234112. DOI: 10.1101/234112. URL: https://www.biorxiv.org/content/early/2017/12/15/234112 (visited on 06/11/2018).

Hinde, Elizabeth, Francesco Cardarelli, and Enrico Gratton (2015). "Spatiotemporal Regulation of Heterochromatin Protein 1- Alpha Oligomerization and Dynamics in Live Cells". In: *Scientific Reports* 5.1. ISSN: 2045-2322. DOI: 10.1038/srep12001. URL: http://www.nature.com/articles/srep12001 (visited on 05/10/2018).

Hipp, Lisa et al. (2019). "Single-Molecule Imaging of the Transcription Factor SRF Reveals Prolonged Chromatin-Binding Kinetics upon Cell Stimulation". In: *Proceedings of the National Academy of Sciences* 116.3, pp. 880–889. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1812734116. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1812734116 (visited on 02/22/2019).

Hnisz, Denes et al. (2017). "A Phase Separation Model for Transcriptional Control". In: *Cell* 169.1. 00000, pp. 13–23. ISSN: 00928674. DOI: 10.1016/j.cell.2017.02.007. URL: http://linkinghub.elsevier.com/retrieve/pii/S009286741730185X (visited on 04/12/2017).

Ho, Han N. et al. (2019). "Identifying Multiple Kinetic Populations of DNA Binding Proteins in Live Cells Using Single-Molecule Fluorescence Imaging". In: *bioRxiv*, p. 509620. DOI: 10.1101/509620. URL: https://www.biorxiv.org/content/early/2019/01/02/509620 (visited on 01/03/2019).

Hoffmann, Marc (2001). "On Estimating the Diffusion Coefficient: Parametric versus Nonparametric". In: *Annales de l'IHP Probabilités et Statistiques*. Vol. 37. 00017, pp. 339–372. URL: http://archive.numdam.org/article/AIHPB_2001__37_3_339_0.pdf (visited on 11/30/2016).

Höfling, Felix and Thomas Franosch (2013). "Anomalous Transport in the Crowded World of Biological Cells". In: *Reports on Progress in Physics* 76.4, p. 046602. ISSN: 0034-4885, 1361-6633. DOI: 10.1088/0034-4885/76/4/046602. URL: http://stacks.iop.org/0034-4885/76/i=4/a=046602?key=crossref.d72b9b76244e9222284e073065248855 (visited on 09/17/2017).

Horlbeck, Max A. et al. (2016). "Nucleosomes Impede Cas9 Access to DNA in Vivo and in Vitro". In: *Elife* 5. 00017, e12677. URL: https://elifesciences.org/content/5/e12677v4 (visited on 10/14/2016).

Howe, Françoise S. et al. (2016). "Is H3K4me3 Instructive for Transcription Activation?" In: *BioEssays* 39.1, e201600095. ISSN: 0265-9247. DOI: 10.1002/bies.201600095. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/bies.201600095 (visited on 04/11/2018).

Hsieh, Tsung-Han S et al. (2016). "Micro-C XL: Assaying Chromosome Conformation from the Nucleosome to the Entire Genome". In: *Nature Methods*. 00001. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4025. URL: http://www.nature.com/doifinder/10.1038/nmeth.4025 (visited on 11/08/2016).

Huang, Bo, Mark Bates, and Xiaowei Zhuang (2009). "Super-Resolution Fluorescence Microscopy". In: *Annual Review of Biochemistry* 78.1. 00749, pp. 993–1016. ISSN: 0066-4154, 1545-4509. DOI: 10.1146/annurev.biochem.77.061906.092014. URL: http://www.annualreviews.org/doi/10.1146/annurev.biochem.77.061906.092014 (visited on 08/23/2016).

Huang, Bo, Wenqin Wang, et al. (2008). "Three-Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy". In: *Science* 319.5864, pp. 810–813. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1153529. pmid: 18174397. URL: http://science.sciencemag.org/content/319/5864/810 (visited on 03/13/2019).

Huang, Ching-Yu et al. (2008). "Dynamic Regulation of C-Myc Proto-Oncogene Expression during Lymphocyte Development Revealed by a GFP-c-Myc Knock-in Mouse". In: *European journal of immunology* 38.2, pp. 342–349.

Huang, Kai, Vadim Backman, and Igal Szleifer (2018). "Interphase Chromatin as a Self-Returning Random Walk: Can DNA Fold into Liquid Trees?" In: DOI: 10.1101/413872. URL: http://biorxiv.org/lookup/doi/10.1101/413872 (visited on 10/15/2018).

Huebinger, Jan et al. (2018). "Quantification of Protein Mobility and Associated Reshuffling of Cytoplasm during Chemical Fixation". In: *bioRxiv*. DOI: 10.1101/410670. URL: http://biorxiv.org/lookup/doi/10.1101/410670 (visited on 11/13/2018).

Hyman, Anthony A., Christoph A. Weber, and Frank Jülicher (2014). "Liquid-Liquid Phase Separation in Biology". In: *Annual Review of Cell and Developmental Biology* 30.1, pp. 39–58. ISSN: 1081-0706. DOI: 10.1146/annurev-cellbio-100913-013325. URL: http://www.annualreviews.org/doi/10.1146/annurev-cellbio-100913-013325 (visited on 11/18/2017).

Ilias, Mohammad and Tom W. Young (2006). "Streptococcus Gordonii Soluble Inorganic Pyrophosphatase: An Important Role for the Interdomain Region in Enzyme Activity". In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1764.7, pp. 1299–1306. ISSN: 15709639. DOI: 10.1016/j.bbapap.2006.05.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S1570963906002056 (visited on 04/25/2019).

Illien, Pierre et al. (2017). "Exothermicity Is Not a Necessary Condition for Enhanced Diffusion of Enzymes". In: *Nano Letters* 17.7, pp. 4415–4420. ISSN: 1530-6984. DOI: 10.1021/acs.nanolett.7b01502. URL: http://dx.doi.org/10.1021/acs.nanolett.7b01502 (visited on 10/13/2017).

Isaacson, S. A., D. M. McQueen, and Charles S. Peskin (2011). "The Influence of Volume Exclusion by Chromatin on the Time Required to Find Specific DNA Binding Sites by Diffusion". In: *Proceedings of the National Academy of Sciences* 108.9, pp. 3815–3820. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1018821108. pmid: 21300894. URL: http://www.pnas.org/content/108/9/3815 (visited on 10/29/2017).

Iwahara, Junji, Markus Zweckstetter, and G. Marius Clore (2006). "NMR Structural and Kinetic Characterization of a Homeodomain Diffusing and Hopping on Nonspecific DNA". In: *Proceedings of the National Academy of Sciences* 103.41, pp. 15062–15067. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0605868103. pmid: 17008406. URL: http://www.pnas.org/content/103/41/15062 (visited on 10/29/2017).

Izeddin, Ignacio, Mohamed El Beheiry, et al. (2012). "PSF Shaping Using Adaptive Optics for Three-Dimensional Single-Molecule Super-Resolution Imaging and Tracking". In: *Optics Express* 20.5, pp. 4957–4967. ISSN: 1094-4087. DOI: 10.1364/OE.20.004957. URL: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-20-5-4957 (visited on 03/13/2019).

Izeddin, Ignacio, Vincent Récamier, et al. (2014). "Single-Molecule Tracking in Live Cells Reveals Distinct Target-Search Strategies of Transcription Factors in the Nucleus". In: *eLife* 3. 00054, e02230. ISSN: 2050-084X. DOI: 10.7554/eLife.02230. pmid: 24925319. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4095940/ (visited on 01/11/2017).

Jain, Ankur and Ronald D. Vale (2017). "RNA Phase Transitions in Repeat Expansion Disorders". In: *Nature* 546.7657. 00002, pp. 243–247. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature22386. URL: http://www.nature.com/doifinder/10.1038/nature22386 (visited on 07/03/2017).

Janicki, Susan M. et al. (2004). "From Silencing to Gene Expression: Real-Time Analysis in Single Cells". In: *Cell* 116.5. 00493, pp. 683–698. URL: http://www.sciencedirect.com/science/article/pii/S0092867404001710 (visited on 08/23/2016).

Jaqaman, Khuloud et al. (2008). "Robust Single-Particle Tracking in Live-Cell Time-Lapse Sequences". In: *Nature Methods* 5.8. 00672, pp. 695–702. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1237. URL: http://www.nature.com/doifinder/10.1038/nmeth.1237 (visited on 03/16/2017).

Jazani, Sina et al. (2018). "An Alternative Framework for Fluorescence Correlation Spectroscopy". In: *bioRxiv*. DOI: 10.1101/426114. URL: http://biorxiv.org/lookup/doi/10.1101/426114 (visited on 11/13/2018).

Jee, Ah-Young et al. (2018). "Enzyme Leaps Fuel Antichemotaxis". In: *Proceedings of the National Academy of Sciences* 115.1, pp. 14–18. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1717844115. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1717844115 (visited on 04/25/2019).

Jeronimo, Célia, Pierre Collin, and François Robert (2016). "The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain". In: *Journal of Molecular Biology* 428.12, pp. 2607–2622. ISSN: 00222836. DOI: 10.1016/j.jmb.2016.02.006. URL: http://linkinghub.elsevier.com/retrieve/pii/S002228361600111X (visited on 11/22/2017).

Jonkers, Iris and John T. Lis (2015). "Getting up to Speed with Transcription Elongation by RNA Polymerase II". In: *Nature Reviews Molecular Cell Biology* 16.3, pp. 167–177. ISSN: 1471-0080. DOI: 10.1038/nrm3953. URL: https://www.nature.com/articles/nrm3953 (visited on 05/28/2018).

Jung, Kwan-Young et al. (2015). "Perturbation of the C-Myc–Max Protein–Protein Interaction via Synthetic α-Helix Mimetics". In: *Journal of Medicinal Chemistry* 58.7, pp. 3002–3024. ISSN: 0022-2623. DOI: 10.1021/jm501440q. URL: https://doi.org/10.1021/jm501440q (visited on 07/03/2018).

Jungmann, Ralf et al. (2016). "Quantitative Super-Resolution Imaging with qPAINT". In: *Nature Methods* 13.5, pp. 439–442. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.3804. URL: http://www.nature.com/articles/nmeth.3804 (visited on 01/17/2018).

Juven-Gershon, T., J.-Y. Hsu, and J. T. Kadonaga (2008). "Caudal, a Key Developmental Regulator, Is a DPE-Specific Transcriptional Factor". In: *Genes & Development* 22.20. 00060, pp. 2823–2830. ISSN: 0890-9369. DOI: 10.1101/gad.1698108. URL: http://genesdev.cshlp.org/cgi/doi/10.1101/gad.1698108 (visited on 12/23/2016).

Juven-Gershon, Tamar, Susan Cheng, and James T. Kadonaga (2006). "Rational Design of a Super Core Promoter That Enhances Gene Expression". In: *Nature Methods* 3.11, pp. 917–922. DOI: DOI:10.1038/NMETH937. URL: http://cyber.sci-hub.cc/MTAuMTAzOC9ubWV0aDkzNw==/juvengershon2006.pdf (visited on 12/23/2016).

Juven-Gershon, Tamar and James T. Kadonaga (2010). "Regulation of Gene Expression via the Core Promoter and the Basal Transcriptional Machinery". In: *Developmental Biology* 339.2. 00295, pp. 225–229. ISSN: 00121606. DOI: 10.1016/j.ydbio.2009.08.009. URL: http://linkinghub.elsevier.com/retrieve/pii/S0012160609011166 (visited on 12/23/2016).

Kadonaga, James T. (2012). "Perspectives on the RNA Polymerase II Core Promoter". In: *Wiley Interdisciplinary Reviews: Developmental Biology* 1.1, pp. 40–51. ISSN: 17597684. DOI: 10.1002/wdev.21. URL: http://doi.wiley.com/10.1002/wdev.21 (visited on 03/28/2016).

Kagey, Michael H. et al. (2010). "Mediator and Cohesin Connect Gene Expression and Chromatin Architecture". In: *Nature* 467.7314, pp. 430–435. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature09380. URL: http://www.nature.com/doifinder/10.1038/nature09380 (visited on 09/22/2017).

Kakui, Yasutaka et al. (2017). "Condensin-Mediated Remodeling of the Mitotic Chromatin Landscape in Fission Yeast". In: *Nature Genetics* 49.10, pp. 1553–1557. ISSN: 1546-1718. DOI: 10.1038/ng.3938. URL: https://www.nature.com/articles/ng.3938 (visited on 04/06/2019).

Kalaskar, Vijay K. and James D. Lauderdale (2014). "Mouse Embryonic Development in a Serum-Free Whole Embryo Culture System". In: *Journal of Visualized Experiments : JoVE* 85. DOI: 10.3791/50803. pmid: 24637443. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4122383/ (visited on 11/07/2018).

Kalhor, Reza et al. (2011). "Genome Architectures Revealed by Tethered Chromosome Conformation Capture and Population-Based Modeling". In: *Nature Biotechnology* 30.1. 00227, pp. 90–98. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.2057. URL: http://www.nature.com/doifinder/10.1038/nbt.2057 (visited on 08/23/2016).

Kalkat, Manpreet et al. (2017). "MYC Deregulation in Primary Human Cancers". In: *Genes* 8.6, p. 151. DOI: 10.3390/genes8060151. URL: https://www.mdpi.com/2073-4425/8/6/151 (visited on 03/30/2019).

Kamar, Ramsey I. et al. (2017). "Facilitated Dissociation of Transcription Factors from Single DNA Binding Sites". In: *Proceedings of the National Academy of Sciences* 114.16. 00000, E3251–E3257. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1701884114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1701884114 (visited on 05/03/2017).

Kao, Shih-Han et al. (2015). "Analysis of Protein Stability by the Cycloheximide Chase Assay". In: *BIO-PROTOCOL* 5.1. ISSN: 2331-8325. DOI: 10.21769/BioProtoc.1374. URL: http://www.bio-protocol.org/e1374 (visited on 10/13/2017).

Karlic, R. et al. (2010). "Histone Modification Levels Are Predictive for Gene Expression". In: *Proceedings of the National Academy of Sciences* 107.7, pp. 2926–2931. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0909344107. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0909344107 (visited on 04/13/2018).

Karrenbauer, Andreas and Dominik Wöll (2013). "Blinking Molecule Tracking". In: *International Symposium on Experimental Algorithms*. 00018. Springer, pp. 308–319. URL: http://link.springer.com/chapter/10.1007/978-3-642-38527-8_28 (visited on 11/10/2016).

Kato, Masato et al. (2012). "Cell-Free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels". In: *Cell* 149.4, pp. 753–767. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2012.

04.017. pmid: 22579281. URL: http://www.cell.com/cell/abstract/S0092-8674(12)00514-4 (visited on 11/22/2017).

Ke, Yuwen et al. (2017). "3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis". In: *Cell* 170.2, 367–381.e20. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2017.06.029. pmid: 28709003, 28709003. URL: https://www.cell.com/cell/abstract/S0092-8674(17)30713-4 (visited on 06/08/2018).

Kieffer-Kwon, Kyong-Rim et al. (2017). "Myc Regulates Chromatin Decompaction and Nuclear Architecture during B Cell Activation". In: *Molecular Cell*. 00000. ISSN: 10972765. DOI: 10.1016/j.molcel.2017.07.013. URL: http://linkinghub.elsevier.com/retrieve/pii/S1097276517305075 (visited on 08/10/2017).

Kim, Jiah et al. (2018). "Nuclear Speckle Fusion via Long-Range Directional Motion Regulates the Number and Size of Speckles". In: *BioRxiv*. DOI: 10.1101/347955. URL: http://biorxiv.org/lookup/doi/10.1101/347955 (visited on 07/24/2018).

Kim, Yongsub et al. (2013). "A Library of TAL Effector Nucleases Spanning the Human Genome". In: *Nature Biotechnology* 31.3, pp. 251–258. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.2517. URL: http://www.nature.com/doifinder/10.1038/nbt.2517 (visited on 08/23/2016).

Kinney, Justin B., Gašper Tkačik, and Curtis G. Callan (2007). "Precise Physical Models of Protein–DNA Interaction from High-Throughput Data". In: *Proceedings of the National Academy of Sciences* 104.2. 00057, pp. 501–506. URL: http://www.pnas.org/content/104/2/501.short (visited on 03/16/2017).

Kiskowski, Maria A., John F. Hancock, and Anne K. Kenworthy (2009). "On the Use of Ripley's K-Function and Its Derivatives to Analyze Domain Size". In: *Biophysical Journal* 97.4. 00084, pp. 1095–1103. ISSN: 00063495. DOI: 10.1016/j.bpj.2009.05.039. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349509010480 (visited on 10/23/2016).

Knight, Spencer C. et al. (2015). "Dynamics of CRISPR-Cas9 Genome Interrogation in Living Cells". In: *Science* 350.6262. 00035, pp. 823–826. URL: http://science.sciencemag.org/content/350/6262/823.short (visited on 10/14/2016).

Knoepfler, Paul S. (2008). "Why Myc? An Unexpected Ingredient in the Stem Cell Cocktail". In: *Cell Stem Cell* 2.1. 00105, pp. 18–21. ISSN: 19345909. DOI: 10.1016/j.stem.2007.12.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S1934590907003207 (visited on 08/23/2016).

Knoepfler, Paul S. et al. (2006). "Myc Influences Global Chromatin Structure". In: 00313.

Kókai, Enikö (2010). "C-Myc Expression in Adult and Embryonic Endothelial Cells". 00000. Universität Ulm. URL: https://doi.org/10.18725/OPARU-1876.

Kolmakov, Kirill et al. (2015). "Far-Red Emitting Fluorescent Dyes for Optical Nanoscopy: Fluorinated Silicon-Rhodamines (SiRF Dyes) and Phosphorylated Oxazines". In: *Chemistry - A European Journal* 21.38, pp. 13344–13356. ISSN: 09476539. DOI: 10.1002/chem.201501394. URL: http://doi.wiley.com/10.1002/chem.201501394 (visited on 12/18/2018).

Konermann, Silvana et al. (2013). "Optical Control of Mammalian Endogenous Transcription and Epigenetic States". In: *Nature*. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature12466. URL: http://www.nature.com/doifinder/10.1038/nature12466 (visited on 09/22/2017).

Kopelman, Raoul (1986). "Rate Processes on Fractals: Theory, Simulations, and Experiments". In: *Journal of Statistical Physics* 42.1, pp. 185–200.

— (1988). "Fractal Reaction Kinetics". In: *Science* 241.4873, pp. 1620–1626. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.241.4873.1620. pmid: 17820893. URL: http://science.sciencemag.org/content/241/4873/1620 (visited on 03/28/2018).

Kornberg, Roger D. and Lubert Stryer (1988). "Statistical Distributions of Nucleosomes: Nonrandom Locations by a Stochastic Mechanism". In: *Nucleic Acids Research* 16.14, pp. 6677–6690. ISSN: 0305-1048. DOI: 10.1093/nar/16.14.6677. URL: https://academic.oup.com/nar/article/16/14/6677/2378544 (visited on 04/06/2019).

Koutelou, Evangelia, Calley L Hirsch, and Sharon YR Dent (2010). "Multiple Faces of the SAGA Complex". In: *Current Opinion in Cell Biology* 22.3. 00127, pp. 374–382. ISSN: 09550674. DOI: 10.1016/j.ceb.2010.03.005. URL: http://linkinghub.elsevier.com/retrieve/pii/S0955067410000293 (visited on 01/13/2017).

Krepelova, Anna et al. (2014). "Myc and Max Genome-Wide Binding Sites Analysis Links the Myc Regulatory Network with the Polycomb and the Core Pluripotency Networks in Mouse Embryonic Stem Cells". In: *PLoS ONE* 9.2. Ed. by Sue Cotterill. 00012, e88933. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0088933. URL: http://dx.plos.org/10.1371/journal.pone.0088933 (visited on 04/20/2017).

Krichevsky, Oleg and Grégoire Bonnet (2002). "Fluorescence Correlation Spectroscopy: The Technique and Its Applications". In: *Reports on Progress in Physics* 65.2, pp. 251–297. ISSN: 0034-4885, 1361-6633. DOI: 10.1088/0034-4885/65/2/203. URL: http://stacks.iop.org/0034-4885/65/i=2/a=203?key=crossref.84cdff198fc426297bdef91ea75099ee (visited on 04/25/2019).

Krog, Jens et al. (2018). "Bayesian Model Selection with Fractional Brownian Motion". In: arXiv: 1804.01365 [cond-mat, physics:physics]. URL: http://arxiv.org/abs/1804.01365 (visited on 06/04/2018).

Kruithof, Maarten et al. (2009). "Single-Molecule Force Spectroscopy Reveals a Highly Compliant Helical Folding for the 30-Nm Chromatin Fiber". In: *Nature Structural & Molecular Biology* 16.5, pp. 534–540. ISSN: 1545-9985. DOI: 10.1038/nsmb.1590. URL: https://www.nature.com/articles/nsmb.1590 (visited on 04/06/2019).

Kues, Thorsten and Ulrich Kubitscheck (2002). "Single Molecule Motion Perpendicular to the Focal Plane of a Microscope: Application to Splicing Factor Dynamics within the Cell Nucleus". In: *Single Molecules* 3.4. 00011, pp. 218–224. ISSN:

1438-5171. DOI: 10.1002/1438-5171(200208)3:4<218::AID-SIMO218>3.0.CO;2-C. URL: http://dx.doi.org/10.1002/1438-5171(200208)3:4%3C218::AID-SIMO218%3E3.0.CO;2-C.

Kujirai, Tomoya et al. (2018). "Structural Basis of the Nucleosome Transition during RNA Polymerase II Passage". In: *Science* 362.6414, pp. 595–598. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aau9904. pmid: 30287617. URL: http://science.sciencemag.org/content/362/6414/595 (visited on 12/04/2018).

Kumar, Niti et al. (2008). "Silencing C-MYC Expression by Targeting Quadruplex in P1 Promoter Using Locked Nucleic Acid Trap". In: *Biochemistry* 47.50, pp. 13179–13188. ISSN: 0006-2960. DOI: 10.1021/bi801064j. URL: https://doi.org/10.1021/bi801064j (visited on 08/24/2018).

Kurland, J. F. and W. P. Tansey (2008). "Myc-Mediated Transcriptional Repression by Recruitment of Histone Deacetylase". In: *Cancer Research* 68.10. 00077, pp. 3624–3629. ISSN: 0008-5472, 1538-7445. DOI: 10.1158/0008-5472.CAN-07-6552. URL: http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-07-6552 (visited on 04/28/2017).

Kwon, Ilmin et al. (2013). "Phosphorylation-Regulated Binding of RNA Polymerase II to Fibrous Polymers of Low-Complexity Domains". In: *Cell* 155.5, pp. 1049–1060. ISSN: 00928674. DOI: 10.1016/j.cell.2013.10.033. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867413013512 (visited on 11/21/2017).

Laine, Romain et al. (2018). "NanoJ: A High-Performance Open-Source Super-Resolution Microscopy Toolbox". In: *BioRxiv*. DOI: 10.1101/432674.

Larson, Adam G. et al. (2017). "Liquid Droplet Formation by HP1α Suggests a Role for Phase Separation in Heterochromatin". In: *Nature* 547.7662, pp. 236–240. ISSN: 1476-4687. DOI: 10.1038/nature22822. URL: https://www.nature.com/articles/nature22822 (visited on 03/26/2018).

Larson, D. R. et al. (2011). "Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene". In: *Science* 332.6028. 00240, pp. 475–478. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1202142. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1202142 (visited on 08/23/2016).

Lawrence, Moyra, Sylvain Daujat, and Robert Schneider (2016). "Lateral Thinking: How Histone Modifications Regulate Gene Expression". In: *Trends in Genetics* 32.1, pp. 42–56. ISSN: 01689525. DOI: 10.1016/j.tig.2015.10.007. URL: http://linkinghub.elsevier.com/retrieve/pii/S0168952515001936 (visited on 04/11/2018).

Lazar-Stefanita, Luciana et al. (2017). "Cohesins and Condensins Orchestrate the 4D Dynamics of Yeast Chromosomes during the Cell Cycle". In: *The EMBO Journal*, e201797342. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.201797342. pmid: 28729434. URL: http://emboj.embopress.org/content/early/2017/07/20/embj.201797342 (visited on 04/06/2019).

Lebedev, D. V. et al. (2008). "Structural Hierarchy of Chromatin in Chicken Erythrocyte Nuclei Based on Small-Angle Neutron Scattering: Fractal Nature of the Large-Scale Chromatin Organization". In: *Crystallography Reports* 53.1, pp. 110–115. ISSN: 1063-7745, 1562-689X. DOI: 10.1134/S1063774508010136. URL: http://link.springer.com/10.1134/S1063774508010136 (visited on 12/05/2017).

Lebedev, D.V. et al. (2005). "Fractal Nature of Chromatin Organization in Interphase Chicken Erythrocyte Nuclei: DNA Structure Exhibits Biphasic Fractal Properties". In: *FEBS Letters* 579.6, pp. 1465–1468. ISSN: 00145793. DOI: 10.1016/j.febslet.2005.01.052. URL: http://doi.wiley.com/10.1016/j.febslet.2005.01.052 (visited on 12/05/2017).

Lee, Alan (2010). "Circular Data". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.4, pp. 477–486. ISSN: 1939-0068. DOI: 10.1002/wics.98. URL: http://onlinelibrary.wiley.com/doi/10.1002/wics.98/abstract (visited on 10/31/2017).

Lee, Antony et al. (2017). "Unraveling the Thousand Word Picture: An Introduction to Super-Resolution Data Analysis". In: *Chemical Reviews* 117.11, pp. 7276–7330. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.6b00729. URL: http://pubs.acs.org/doi/10.1021/acs.chemrev.6b00729 (visited on 12/11/2017).

Leemans, Christ et al. (2018). "Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in Lamina-Associated Domains". In: *bioRxiv*, p. 464081. DOI: 10.1101/464081. URL: https://www.biorxiv.org/content/10.1101/464081v1 (visited on 03/05/2019).

Leong, Philip HW et al. (2005). "A Comment on the Implementation of the Ziggurat Method". In: *Journal of Statistical Software* 12.7, pp. 1–4.

Levens, D. (2008). "How the C-Myc Promoter Works and Why It Sometimes Does Not". In: *JNCI Monographs* 2008.39. 00029, pp. 41–43. ISSN: 1052-6773, 1745-6614. DOI: 10.1093/jncimonographs/lgn004. URL: http://jncimono.oxfordjournals.org/cgi/doi/10.1093/jncimonographs/lgn004 (visited on 08/23/2016).

Levet, Florian et al. (2015). "SR-Tesseler: A Method to Segment and Quantify Localization-Based Super-Resolution Microscopy Data". In: *Nature Methods* 12.11, pp. 1065–1071. ISSN: 1548-7091. DOI: 10.1038/nmeth.3579. URL: http://www.nature.com/nmeth/journal/v12/n11/full/nmeth.3579.html (visited on 10/01/2017).

Levine, Michael, Claudia Cattoglio, and Robert Tjian (2014). "Looping Back to Leap Forward: Transcription Enters a New Era". In: *Cell* 157.1, pp. 13–25. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2014.02.009. pmid: 24679523. URL: https://www.cell.com/cell/abstract/S0092-8674(14)00201-3 (visited on 06/08/2018).

Li, Haoyan et al. (2018). "Inhibition of cIAP1 as a Strategy for Targeting C-MYC–Driven Oncogenic Activity". In: *Proceedings of the National Academy of Sciences* 115.40, E9317–E9324. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1807711115. pmid: 30181285. URL: http://www.pnas.org/content/115/40/E9317 (visited on 10/03/2018).

Li, Pilong et al. (2012). "Phase Transitions in the Assembly of Multivalent Signalling Proteins". In: *Nature* 483.7389, pp. 336–340. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature10879. URL: http://www.nature.com/doifinder/10.1038/nature10879 (visited on 11/18/2017).

Li, Yiming et al. (2017). "Fast, Robust and Precise 3D Localization for Arbitrary Point Spread Functions". In: *bioRxiv*. 00000. DOI: 10.1101/172643. URL: http://biorxiv.org/content/early/2017/08/08/172643.abstract.

Lieberman-Aiden, E. et al. (2009). "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome". In: *Science* 326.5950, pp. 289–293. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1181369. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1181369 (visited on 11/22/2017).

Lim, Bomyi et al. (2018). "Visualization of Transvection in Living Drosophila Embryos". In: *Molecular Cell* 70.2, 287–296.e6. ISSN: 10972765. DOI: 10.1016/j.molcel.2018.02.029. URL: http://linkinghub.elsevier.com/retrieve/pii/S1097276518301746 (visited on 06/22/2018).

Lim, Wooi F. et al. (2016). "Directing an Artificial Zinc Finger Protein to New Targets by Fusion to a Non-DNA-Binding Domain". In: *Nucleic Acids Research* 44.7. 00003, pp. 3118–3130. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkv1380. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1380 (visited on 04/11/2017).

Limouse, Charles et al. (2017). "Intramolecular Dynamics of Single Molecules in Free Diffusion". In: *bioRxiv*. 00000, p. 120311. URL: http://biorxiv.org/content/early/2017/03/24/120311.abstract (visited on 04/02/2017).

Lin, Charles Y. et al. (2012). "Transcriptional Amplification in Tumor Cells with Elevated C-Myc". In: *Cell* 151.1. 00000, pp. 56–67. ISSN: 00928674. DOI: 10.1016/j.cell.2012.08.026. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867412010574 (visited on 08/23/2016).

Lindahl, Erik, Berk Hess, and David van der Spoel (2001). "GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis". In: *Journal of Molecular Modeling* 7.8, pp. 306–317. ISSN: 1610-2940, 0948-5023. DOI: 10.1007/s008940100045. URL: http://link.springer.com/10.1007/s008940100045 (visited on 02/20/2018).

Lindén, Martin, Vladimir Ćurić, Elias Amselem, et al. (2017). "Pointwise Error Estimates in Localization Microscopy". In: *Nature Communications* 8. 00000, p. 15115. ISSN: 2041-1723. DOI: 10.1038/ncomms15115. URL: http://www.nature.com/doifinder/10.1038/ncomms15115 (visited on 07/17/2017).

Lindén, Martin, Vladimir Ćurić, Alexis Boucharin, et al. (2016). "Simulated Single Molecule Microscopy with SMeagol". In: *Bioinformatics* 32.15. 00002, pp. 2394–2395. ISSN: 1367-4803, 1460-2059. DOI: 10.1093/bioinformatics/btw109. URL: http://bioinformatics.oxfordjournals.org/lookup/doi/10.1093/bioinformatics/btw109 (visited on 11/17/2016).

Lindén, Martin and Johan Elf (2018). "Variational Algorithms for Analyzing Noisy Multistate Diffusion Trajectories". In: *Biophysical Journal* 115.2, pp. 276–282. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2018.05.027. URL: http://www.sciencedirect.com/science/article/pii/S0006349518306659 (visited on 07/18/2018).

Liu, D. S. et al. (2014). "Computational Design of a Red Fluorophore Ligase for Site-Specific Protein Labeling in Living Cells". In: *Proceedings of the National Academy of Sciences* 111.43, E4551–E4559. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1404736111. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1404736111 (visited on 09/25/2017).

Liu, Hui et al. (2017). "Stochastic Protein Labeling Enables Long-Term Single Molecule Observation In Vivo". In: *bioRxiv*. 00000, p. 116186. URL: http://biorxiv.org/content/early/2017/03/13/116186.abstract (visited on 04/02/2017).

Liu, J and D. Levens (2006). "Making Myc". In: *CTMI*. 00090.

Liu, Jiangang et al. (2006). "Intrinsic Disorder in Transcription Factors". In: *Biochemistry* 45.22, pp. 6873–6888. ISSN: 0006-2960. DOI: 10.1021/bi0602718. URL: http://dx.doi.org/10.1021/bi0602718 (visited on 12/06/2017).

Liu, Zhe, Luke?D. Lavis, and Eric Betzig (2015). "Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level". In: *Molecular Cell* 58.4. 00085, pp. 644–659. ISSN: 10972765. DOI: 10.1016/j.molcel.2015.02.033. URL: http://linkinghub.elsevier.com/retrieve/pii/S1097276515001653 (visited on 06/22/2017).

Liu, Zhe, Wesley R. Legant, et al. (2014). "3D Imaging of Sox2 Enhancer Clusters in Embryonic Stem Cells". In: *Elife* 3. 00029, e04236. URL: https://elifesciences.org/content/3/e04236v1 (visited on 10/14/2016).

Loffreda, Alessia et al. (2017). "Live-Cell P53 Single-Molecule Binding Is Modulated by C-Terminal Acetylation and Correlates with Transcriptional Activity". In: *Nature Communications* 8.1, p. 313. ISSN: 2041-1723. DOI: 10.1038/s41467-017-00398-7. URL: https://www.nature.com/articles/s41467-017-00398-7 (visited on 11/05/2017).

Lorén, Niklas et al. (2015). "Fluorescence Recovery after Photobleaching in Material and Life Sciences: Putting Theory into Practice". In: *Quarterly Reviews of Biophysics* 48.3, pp. 323–387. ISSN: 0033-5835, 1469-8994. DOI: 10.1017/S0033583515000013. URL: https://www.cambridge.org/core/journals/quarterly-reviews-of-biophysics/article/fluorescence-recovery-after-photobleaching-in-material-and-life-sciences-putting-theory-into-practice/F9C82E9F20B775957AA6468359A59B89 (visited on 10/27/2017).

Lorenzin, Francesca et al. (2016). "Different Promoter Affinities Account for Specificity in MYC-Dependent Gene Regulation". In: *eLife* 5, e15161. ISSN: 2050-084X. DOI: 10.7554/eLife.15161. URL: https://elifesciences.org/articles/15161 (visited on 05/28/2018).

Los, Georgyi V. et al. (2008). "HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis". In: *ACS Chemical Biology* 3.6. 00593, pp. 373–382. ISSN: 1554-8929, 1554-8937. DOI: 10.1021/cb800025k. URL: http://pubs.acs.org/doi/abs/10.1021/cb800025k (visited on 07/28/2016).

Louder, Robert K. et al. (2016). "Structure of Promoter-Bound TFIID and Model of Human Pre-Initiation Complex Assembly". In: *Nature* 531.7596, pp. 604–609. ISSN: 0028-0836. DOI: 10.1038/nature17394. URL: http://www.nature.com/nature/journal/v531/n7596/full/nature17394.html (visited on 09/26/2017).

Lovinger, Justin (2018). *Optimal: A Python Metaheuristic Optimization Library. Currently Supports Genetic Algorithms, Gravitational Search, and Cross Entropy.* Version 0.2.0. URL: https://github.com/JustinLovinger/Optimal (visited on 02/28/2018).

Lu, Xiaodong et al. (2016). "Multiple P-TEFbs Cooperatively Regulate the Release of Promoter-Proximally Paused RNA Polymerase II". In: *Nucleic Acids Research* 44.14, pp. 6853–6867. ISSN: 0305-1048. DOI: 10.1093/nar/gkw571. URL: https://academic.oup.com/nar/article/44/14/6853/2468206 (visited on 01/09/2018).

Lucas, Joseph S. et al. (2014). "3D Trajectories Adopted by Coding and Regulatory DNA Elements: First-Passage Times for Genomic Interactions". In: *Cell* 158.2. 00041, pp. 339–352. ISSN: 00928674. DOI: 10.1016/j.cell.2014.05.036. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867414007375 (visited on 12/23/2016).

Luger, Karolin et al. (1997). "Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution". In: *Nature* 389.6648, pp. 251–260. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/38444. URL: http://www.nature.com/articles/38444 (visited on 04/07/2019).

Lupiáñez, Darío G. et al. (2015). "Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions". In: *Cell* 161.5, pp. 1012–1025. ISSN: 00928674. DOI: 10.1016/j.cell.2015.04.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867415003773 (visited on 02/26/2016).

Lüscher, Bernhard and Jörg Vervoorts (2012). "Regulation of Gene Transcription by the Oncoprotein MYC". In: *Gene* 494.2. 00079, pp. 145–160. ISSN: 03781119. DOI: 10.1016/j.gene.2011.12.027. URL: http://linkinghub.elsevier.com/retrieve/pii/S0378111911008110 (visited on 08/23/2016).

Ma, Wenxiu and Wing Hung Wong (2011). "The Analysis of ChIP-Seq Data". In: *Methods in Enzymology*. Vol. 497. 00018. Elsevier, pp. 51–73. ISBN: 978-0-12-385075-1. URL: http://linkinghub.elsevier.com/retrieve/pii/B9780123850751000032 (visited on 08/23/2016).

Machado, Sarah, Vincent Mercier, and Nicolas Chiaruttini (2018). "LimeSeg: A Coarsed-Grained Lipid Membrane Simulation for 3D Image Segmentation". In: *bioRxiv*, p. 267534. DOI: 10.1101/267534. URL: https://www.biorxiv.org/content/early/2018/02/18/267534 (visited on 02/19/2018).

Mack, Andrew H. et al. (2012). "Kinetics and Thermodynamics of Phenotype: Unwinding and Rewinding the Nucleosome". In: *Journal of Molecular Biology* 423.5, pp. 687–701. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2012.08.021. URL: http://www.sciencedirect.com/science/article/pii/S0022283612006985 (visited on 04/06/2019).

Magdziarz, Marcin et al. (2009). "Fractional Brownian Motion Versus the Continuous-Time Random Walk: A Simple Test for Subdiffusive Dynamics". In: *Physical Review Letters* 103.18. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.103.180602. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.180602 (visited on 03/05/2018).

Maldonado, Andres Y., David S. Burz, and Alexander Shekhtman (2011). "In-Cell NMR Spectroscopy". In: *Progress in nuclear magnetic resonance spectroscopy* 59.3, pp. 197–212. ISSN: 0079-6565. DOI: 10.1016/j.pnmrs.2010.11.002. pmid: 21920217. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3175053/ (visited on 10/30/2017).

Manley, Suliana et al. (2008). "High-Density Mapping of Single-Molecule Trajectories with Photoactivated Localization Microscopy". In: *Nature Methods* 5.2. 00666, pp. 155–157. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1176. URL: http://www.nature.com/doifinder/10.1038/nmeth.1176 (visited on 07/01/2017).

Manolov, George and Yanka Manolova (1972). "Marker Band in One Chromosome 14 from Burkitt Lymphomas". In: *Nature* 237.5349, p. 33. ISSN: 1476-4687. DOI: 10.1038/237033a0. URL: https://www.nature.com/articles/237033a0 (visited on 03/30/2019).

Marfil, V. et al. (2015). "Growth-Promoting and Tumourigenic Activity of c-Myc Is Suppressed by Hhex". In: *Oncogene* 34.23, pp. 3011–3022. ISSN: 1476-5594. DOI: 10.1038/onc.2014.240. URL: https://www.nature.com/articles/onc2014240 (visited on 07/03/2018).

Margineanu, Anca et al. (2016). "Screening for Protein-Protein Interactions Using Förster Resonance Energy Transfer (FRET) and Fluorescence Lifetime Imaging Microscopy (FLIM)". In: *Scientific Reports* 6.1. ISSN: 2045-2322. DOI: 10.1038/srep28186. URL: http://www.nature.com/articles/srep28186 (visited on 03/26/2018).

Marklund, Emil et al. (2018). "Direct Observation of Rotation-Coupled Protein Diffusion along DNA on the Microsecond Timescale". In: *bioRxiv*, p. 401414. DOI: 10.1101/401414. URL: https://www.biorxiv.org/content/early/2018/08/27/401414 (visited on 08/28/2018).

Marsh, Joseph A. and Julie D. Forman-Kay (2010). "Sequence Determinants of Compaction in Intrinsically Disordered Proteins". In: *Biophysical journal* 98.10, pp. 2383–2390.

Martell, Jeffrey D. et al. (2016). "A Split Horseradish Peroxidase for the Detection of Intercellular Protein-Protein Interactions and Sensitive Visualization of Synapses". In: *Nature Biotechnology* 34.7. 00000, pp. 774–780. ISSN: 1087-0156. DOI: 10.1038/nbt.3563. URL: http://www.nature.com/nbt/journal/v34/n7/full/nbt.3563.html (visited on 08/25/2016).

Martens, Koen J.A. et al. (2018). "An Open Microscopy Framework Suited for Tracking dCas9 in Live Bacteria". In: *BioRxiv*. DOI: 10.1101/437137. URL: http://biorxiv.org/lookup/doi/10.1101/437137 (visited on 11/29/2018).

Martin, Brent R et al. (2005). "Mammalian Cell–Based Optimization of the Biarsenical-Binding Tetracysteine Motif for Improved Fluorescence and Affinity". In: *Nature Biotechnology* 23.10, pp. 1308–1314. ISSN: 1087-0156. DOI: 10.1038/nbt1136. URL: http://www.nature.com/doifinder/10.1038/nbt1136 (visited on 08/23/2016).

Masson, J.-B. et al. (2009). "Inferring Maps of Forces inside Cell Membrane Microdomains". In: *Physical Review Letters* 102.4. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.102.048103. URL: https://link.aps.org/doi/10.1103/PhysRevLett.102.048103 (visited on 12/05/2017).

Masson, Jean-Baptiste et al. (2014). "Mapping the Energy and Diffusion Landscapes of Membrane Proteins at the Cell Surface Using High-Density Single-Molecule Imaging and Bayesian Inference: Application to the Multiscale Dynamics of Glycine Receptors in the Neuronal Membrane". In: *Biophysical Journal* 106.1. 00039, pp. 74–83. ISSN: 00063495. DOI: 10.1016/j.bpj.2013.10.027. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349513011946 (visited on 08/28/2016).

Mata, João F. et al. (2012). "A Rapid FACS-Based Strategy to Isolate Human Gene Knockin and Knockout Clones". In: *PLoS ONE* 7.2. Ed. by Beth A. Sullivan. 00009, e32646. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0032646. URL: http://dx.plos.org/10.1371/journal.pone.0032646 (visited on 08/23/2016).

Mateo, Leslie J. et al. (2019). "Visualizing DNA Folding and RNA in Embryos at Single-Cell Resolution". In: *Nature*, p. 1. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1035-4. URL: https://www.nature.com/articles/s41586-019-1035-4 (visited on 03/27/2019).

Matera, A. Gregory (1999). "Nuclear Bodies: Multifaceted Subdomains of the Interchromatin Space". In: *Trends in Cell Biology* 9.8, pp. 302–309. ISSN: 0962-8924. DOI: 10.1016/S0962-8924(99)01606-2. URL: http://www.sciencedirect.com/science/article/pii/S0962892499016062 (visited on 04/06/2019).

Matsuda, Hiroaki et al. (2014). "Macromolecular Crowding as a Regulator of Gene Transcription". In: *Biophysical Journal* 106.8, pp. 1801–1810. ISSN: 00063495. DOI: 10.1016/j.bpj.2014.02.019. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349514002252 (visited on 09/29/2017).

Matsuoka, Satomi, Tatsuo Shibata, and Masahiro Ueda (2009). "Statistical Analysis of Lateral Diffusion and Multistate Kinetics in Single-Molecule Imaging". In: *Biophysical Journal* 97.4. 00025, pp. 1115–1124. ISSN: 00063495. DOI: 10.1016/j.bpj.2009.06.007. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349509011060 (visited on 12/01/2016).

Matysik, Artur and Rachel S. Kraut (2014). "TrackArt: The User Friendly Interface for Single Molecule Tracking Data Analysis and Simulation Applied to Complex Diffusion in Mica Supported Lipid Bilayers". In: *BMC research notes* 7.1. 00012, p. 274. URL: https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-7-274 (visited on 06/18/2017).

Mazza, Davide, Sourav Ganguly, and James G. McNally (2013). "Monitoring Dynamic Binding of Chromatin Proteins In Vivo by Single-Molecule Tracking". In: *Imaging Gene Expression*. Ed. by Yaron Shav-Tal. Vol. 1042. 00000. Totowa, NJ: Humana Press, pp. 117–137. ISBN: 978-1-62703-525-5 978-1-62703-526-2. DOI: 10.1007/978-1-62703-526-2_9. URL: http://link.springer.com/10.1007/978-1-62703-526-2_9 (visited on 05/09/2017).

Mazza, Davide, Florian Mueller, et al. (2013). "Convergence of Chromatin Binding Estimates in Live Cells". In: *Nat Meth* 10.8. 00006, pp. 691–692. ISSN: 1548-7091. URL: http://dx.doi.org/10.1038/nmeth.2573.

Mazza, D. et al. (2012). "A Benchmark for Chromatin Binding Measurements in Live Cells". In: *Nucleic Acids Research* 40.15. 00072, e119–e119. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gks701. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks701 (visited on 11/30/2016).

McDonald, Kent L. et al. (2007). "Recent Advances in High-Pressure Freezing". In: *Electron Microscopy*. Ed. by John Kuo. Red. by John M. Walker. Vol. 369. Totowa, NJ: Humana Press, pp. 143–173. ISBN: 978-1-58829-573-6 978-1-59745-294-6. DOI: 10.1007/978-1-59745-294-6_8. URL: http://link.springer.com/10.1007/978-1-59745-294-6_8 (visited on 08/26/2018).

McGuffee, Sean R. and Adrian H. Elcock (2010). "Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm". In: *PLoS Computational Biology* 6.3. Ed. by James M. Briggs, e1000694. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000694. URL: http://dx.plos.org/10.1371/journal.pcbi.1000694 (visited on 09/22/2017).

McHale, Kevin, Andrew J. Berglund, and Hideo Mabuchi (2004). "Bayesian Estimation for Species Identification in Single-Molecule Fluorescence Microscopy". In: *Biophysical Journal* 86.6, pp. 3409–3422. ISSN: 00063495. DOI: 10.1529/biophysj.103.038414. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349504743878 (visited on 09/25/2017).

McMahon, Steven B., Marcelo A. Wood, and Michael D. Cole (2000). "The Essential Cofactor TRRAP Recruits the Histone Acetyltransferase hGCN5 to C-Myc". In: *Molecular and cellular biology* 20.2. 00427, pp. 556–562. URL: http://mcb.asm.org/content/20/2/556.short (visited on 08/25/2016).

McSwiggen, David Trombley et al. (2018). "Transient DNA Binding Induces RNA Polymerase II Compartmentalization During Herpesviral Infection Distinct From Phase Separation". In: *bioRxiv*, p. 375071. DOI: 10.1101/375071. URL: https://www.biorxiv.org/content/early/2018/07/23/375071.1 (visited on 07/24/2018).

Medh, Rheem D. et al. (2001). "Constitutive Expression of Ectopic C-Myc Delays Glucocorticoid-Evoked Apoptosis of Human Leukemic CEM-C7 Cell". In: *Oncogene* 20. 00000, pp. 4629–4639.

Menssen, Antje and Heiko Hermeking (2002). "Characterization of the C-MYC-Regulated Transcriptome by SAGE: Identification and Analysis of c-MYC Target Genes". In: *Proceedings of the National Academy of Sciences* 99.9. 00362, pp. 6274–6279. URL: http://www.pnas.org/content/99/9/6274.short (visited on 08/25/2016).

Merkenschlager, Matthias and Elphège P. Nora (2016). "CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation". In: *Annual Review of Genomics and Human Genetics* 17.1, pp. 17–43. ISSN: 1527-8204, 1545-293X. DOI:

10.1146/annurev-genom-083115-022339. URL: http://www.annualreviews.org/doi/10.1146/annurev-genom-083115-022339 (visited on 11/11/2017).

Metzler, Ralf, Jae-Hyung Jeon, et al. (2014). "Anomalous Diffusion Models and Their Properties: Non-Stationarity, Non-Ergodicity, and Ageing at the Centenary of Single Particle Tracking". In: *Phys. Chem. Chem. Phys.* 16.44, pp. 24128–24164. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/C4CP03465A. URL: http://xlink.rsc.org/?DOI=C4CP03465A (visited on 11/03/2017).

Metzler, Ralf, Vincent Tejedor, et al. (2009). "Analysis of Single Particle Trajectories: From Normal to Anomalous Diffusion." In: *Acta Physica Polonica B* 40.5.

Meyer, B. et al. (2012). "Geometry-Induced Bursting Dynamics in Gene Expression". In: *Biophysical Journal* 102.9. 00023, pp. 2186–2191. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.03.060. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349512004018 (visited on 04/26/2017).

Michalet, Xavier (2010). "Mean Square Displacement Analysis of Single-Particle Trajectories with Localization Error: Brownian Motion in an Isotropic Medium". In: *Physical Review E* 82.4. 00224. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.82.041914. URL: https://link.aps.org/doi/10.1103/PhysRevE.82.041914 (visited on 06/27/2017).

Michalet, Xavier and Andrew J. Berglund (2012). "Optimal Diffusion Coefficient Estimation in Single-Particle Tracking". In: *Physical Review E* 85.6. 00076. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.85.061916. URL: http://link.aps.org/doi/10.1103/PhysRevE.85.061916 (visited on 08/28/2016).

Miller, Helen et al. (2017). "Ultra-Fast Super-Resolution Imaging of Biomolecular Mobility in Tissues". In: *bioRxiv*. 00000. DOI: 10.1101/179747. URL: http://biorxiv.org/content/early/2017/08/23/179747.abstract.

Milo, Ron and Rob Phillips (2016). *Cell Biology by the Numbers*. New York, NY: Garland Science, Taylor & Francis Group. 356 pp. ISBN: 978-0-8153-4537-4.

Mirny, Leonid A. (2010). "Nucleosome-Mediated Cooperativity between Transcription Factors". In: *Proceedings of the National Academy of Sciences* 107.52. 00144, pp. 22534–22539. URL: http://www.pnas.org/content/107/52/22534.short (visited on 02/13/2017).

Miron, E. et al. (2019). "Chromatin Arranges in Filaments of Blobs with Nanoscale Functional Zonation". In: *bioRxiv*. DOI: 10.1101/566638. URL: http://biorxiv.org/lookup/doi/10.1101/566638 (visited on 03/27/2019).

Mlodzianoski, Michael J. et al. (2011). "Sample Drift Correction in 3D Fluorescence Photoactivation Localization Microscopy". In: *Optics express* 19.16. 00115, pp. 15009–15019. URL: https://www.osapublishing.org/abstract.cfm?uri=oe-19-16-15009 (visited on 07/21/2017).

Moler, Cleve and Charles Van Loan (2003). "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later". In: *SIAM Review* 45.1, pp. 3–49. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/S00361445024180. URL: http://epubs.siam.org/doi/10.1137/S00361445024180 (visited on 10/27/2017).

Molliex, Amandine et al. (2015). "Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization". In: *Cell* 163.1, pp. 123–133. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2015.09.015. pmid: 26406374,26406374. URL: http://www.cell.com/cell/abstract/S0092-8674(15)01176-9 (visited on 11/18/2017).

Monnier, Nilah, Zachary Barry, et al. (2015). "Inferring Transient Particle Transport Dynamics in Live Cells". In: *Nature Methods* 12.9. 00011, pp. 838–840. ISSN: 1548-7091. DOI: 10.1038/nmeth.3483. URL: http://www.nature.com/nmeth/journal/v12/n9/full/nmeth.3483.html (visited on 10/19/2016).

Monnier, Nilah, Syuan-Ming Guo, et al. (2012). "Bayesian Approach to MSD-Based Analysis of Particle Motion in Live Cells". In: *Biophysical Journal* 103.3, pp. 616–626. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.06.029. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349512007187 (visited on 09/25/2017).

Montavon, Thomas et al. (2011). "A Regulatory Archipelago Controls Hox Genes Transcription in Digits". In: *Cell* 147.5. 00249, pp. 1132–1145. ISSN: 00928674. DOI: 10.1016/j.cell.2011.10.023. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867411012736 (visited on 03/16/2017).

Morf, Jorg et al. (2017). "Spatial RNA Proximities Reveal a Bipartite Nuclear Transcriptome and Territories of Differential Density and Transcription Elongation Rates". In: *bioRxiv*, p. 196147. DOI: 10.1101/196147. URL: https://www.biorxiv.org/content/early/2017/09/29/196147 (visited on 10/02/2017).

Morisaki, Tatsuya et al. (2014). "Single-Molecule Analysis of Transcription Factor Binding at Transcription Sites in Live Cells". In: *Nature Communications* 5. 00044. ISSN: 2041-1723. DOI: 10.1038/ncomms5456. URL: http://www.nature.com/doifinder/10.1038/ncomms5456 (visited on 05/09/2017).

Morris, Stephanie A et al. (2013). "Overlapping Chromatin-Remodeling Systems Collaborate Genome Wide at Dynamic Chromatin Transitions". In: *Nature Structural & Molecular Biology* 21.1. 00061, pp. 73–81. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/nsmb.2718. URL: http://www.nature.com/doifinder/10.1038/nsmb.2718 (visited on 07/31/2017).

Morrison, Alexander H. et al. (2012). "Quantifying the Bicoid Morphogen Gradient in Living Fly Embryos". In: *Cold Spring Harbor Protocols* 2012.4, pdb.top068536. ISSN: 1940-3402, 1559-6095. DOI: 10.1101/pdb.top068536. pmid: 22474658. URL: http://cshprotocols.cshlp.org/content/2012/4/pdb.top068536 (visited on 04/05/2019).

Mortensen, Kim I et al. (2010). "Optimized Localization Analysis for Single-Molecule Tracking and Super-Resolution Microscopy". In: *Nature Methods* 7.5, pp. 377–381. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1447. URL: http://www.nature.com/doifinder/10.1038/nmeth.1447 (visited on 09/25/2017).

Muddana, Hari S. et al. (2010). "Substrate Catalysis Enhances Single-Enzyme Diffusion". In: *Journal of the American Chemical Society* 132.7, pp. 2110–2111. ISSN: 0002-7863. DOI: 10.1021/ja908773a. URL: http://dx.doi.org/10.1021/ja908773a (visited on 10/29/2017).

Mueller, Florian, Davide Mazza, et al. (2010). "FRAP and Kinetic Modeling in the Analysis of Nuclear Protein Dynamics: What Do We Really Know?" In: *Current Opinion in Cell Biology* 22.3. 00129, pp. 403–411. ISSN: 09550674. DOI: 10.1016/j.ceb.2010.03.002. URL: http://linkinghub.elsevier.com/retrieve/pii/S0955067410000268 (visited on 06/29/2017).

Mueller, Florian, Timothy J. Stasevich, et al. (2013). "Quantifying Transcription Factor Kinetics: At Work or at Play?" In: *Critical Reviews in Biochemistry and Molecular Biology* 48.5. 00038, pp. 492–514. ISSN: 1040-9238, 1549-7798. DOI: 10.3109/10409238.2013.833891. URL: http://www.tandfonline.com/doi/full/10.3109/10409238.2013.833891 (visited on 08/01/2017).

Mueller, Florian, Paul Wach, and James G. McNally (2008). "Evidence for a Common Mode of Transcription Factor Interaction with Chromatin as Revealed by Improved Quantitative Fluorescence Recovery after Photobleaching". In: *Biophysical Journal* 94.8. 00139, pp. 3323–3339. ISSN: 00063495. DOI: 10.1529/biophysj.107.123182. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349508704880 (visited on 04/03/2017).

Muhar, Matthias et al. (2018). "SLAM-Seq Defines Direct Gene-Regulatory Functions of the BRD4-MYC Axis". In: *Science* 360.6390, pp. 800–805. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aao2793. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aao2793 (visited on 08/28/2018).

Murakami, Kenji et al. (2015). "Structure of an RNA Polymerase II Preinitiation Complex". In: *Proceedings of the National Academy of Sciences* 112.44, pp. 13543–13548. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1518255112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1518255112 (visited on 11/21/2017).

Murayama, Yasuto, Catarina P. Samora, et al. (2018). "Establishment of DNA-DNA Interactions by the Cohesin Ring". In: *Cell* 172.3, 465–477.e15. ISSN: 00928674. DOI: 10.1016/j.cell.2017.12.021. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867417315015 (visited on 04/06/2019).

Murayama, Yasuto and Frank Uhlmann (2014). "Biochemical Reconstitution of Topological DNA Binding by the Cohesin Ring". In: *Nature* 505.7483, pp. 367–371. ISSN: 1476-4687. DOI: 10.1038/nature12867. URL: https://www.nature.com/articles/nature12867 (visited on 04/06/2019).

— (2015). "DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism". In: *Cell* 163.7, pp. 1628–1640. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2015.11.030. pmid: 26687354. URL: https://www.cell.com/cell/abstract/S0092-8674(15)01549-4 (visited on 04/06/2019).

Mussolino, C. et al. (2014). "TALENs Facilitate Targeted Genome Editing in Human Cells with High Specificity and Low Cytotoxicity". In: *Nucleic Acids Research* 42.10, pp. 6762–6773. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gku305. URL: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku305 (visited on 08/23/2016).

Myant, Kevin et al. (2015). "Serine 62-Phosphorylated MYC Associates with Nuclear Lamins and Its Regulation by CIP2A Is Essential for Regenerative Proliferation". In: *Cell Reports* 12.6, pp. 1019–1031. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2015.07.003. URL: http://www.sciencedirect.com/science/article/pii/S2211124715007299 (visited on 05/28/2018).

Nadarajah, Saralees and Tibor K. Pogány (2016). "On the Distribution of the Product of Correlated Normal Random Variables". In: *Comptes Rendus Mathematique* 354.2, pp. 201–204. ISSN: 1631073X. DOI: 10.1016/j.crma.2015.10.019. URL: http://linkinghub.elsevier.com/retrieve/pii/S1631073X15002873 (visited on 10/29/2017).

Nagano, Takashi, Yaniv Lubling, Tim J. Stevens, et al. (2013). "Single-Cell Hi-C Reveals Cell-to-Cell Variability in Chromosome Structure". In: *Nature* 502.7469. 00304, pp. 59–64. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature12593. URL: http://www.nature.com/doifinder/10.1038/nature12593 (visited on 08/20/2016).

Nagano, Takashi, Yaniv Lubling, Csilla Várnai, et al. (2017). "Cell-Cycle Dynamics of Chromosomal Organization at Single-Cell Resolution". In: *Nature* 547.7661, pp. 61–67. ISSN: 1476-4687. DOI: 10.1038/nature23001. URL: https://www.nature.com/articles/nature23001 (visited on 04/06/2019).

Nair, Satish K. and Stephen K. Burley (2003). "X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA: Molecular Bases of Regulation by Proto-Oncogenic Transcription Factors". In: *Cell* 112.2. 00000, pp. 193–205. URL: http://www.sciencedirect.com/science/article/pii/S0092867402012849 (visited on 08/23/2016).

Narlikar, Geeta J., Ramasubramanian Sundaramoorthy, and Tom Owen-Hughes (2013). "Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes". In: *Cell* 154.3, pp. 490–503. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.07.011. URL: http://www.sciencedirect.com/science/article/pii/S0092867413008829 (visited on 04/06/2019).

Nasmyth, Kim (2001). "Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis". In: *Annual Review of Genetics* 35.1, pp. 673–745. DOI: 10.1146/annurev.genet.35.102401.091334. pmid: 11700297. URL: https://doi.org/10.1146/annurev.genet.35.102401.091334 (visited on 04/06/2019).

Naumova, Natalia et al. (2013). "Organization of the Mitotic Chromosome". In: *Science* 342.6161, pp. 948–953. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1236083. pmid: 24200812. URL: http://science.sciencemag.org/content/342/6161/948 (visited on 04/06/2019).

Nehme, Elias, Eran Hershko, et al. (2019). "Deep Learning for Dense and Multicolor Localization Microscopy (Conference Presentation)". In: *Single Molecule Spectroscopy and Superresolution Imaging XII*. Single Molecule Spectroscopy and Superresolution Imaging XII. Vol. 10884. International Society for Optics and Photonics, 108840R. DOI: 10.1117/12.2506499. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10884/108840R/Deep-

learning-for-dense-and-multicolor-localization-microscopy-Conference-Presentation/10.1117/12.2506499.short (visited on 04/24/2019).

Nehme, Elias, Lucien E. Weiss, et al. (2018). "Deep-STORM: Super-Resolution Single-Molecule Microscopy by Deep Learning". In: *Optica* 5.4, pp. 458–464. ISSN: 2334-2536. DOI: 10.1364/OPTICA.5.000458. URL: https://www.osapublishing.org/optica/abstract.cfm?uri=optica-5-4-458 (visited on 04/24/2019).

Nesbit, Chadd E., Jean M. Tersak, and Edward V. Prochownik (1999). "MYC Oncogenes and Human Neoplastic Disease." In: *Oncogene* 18.19.

nickmayorov (2017). *Basic Algorithms for Nonlinear Least Squares*. 00000. URL: https://nmayorov.wordpress.com/2015/06/18/basic-algorithms-for-nonlinear-least-squares/ (visited on 04/29/2017).

Nicolas, Damien et al. (2018). "Modulation of Transcriptional Burst Frequency by Histone Acetylation". In: *Proceedings of the National Academy of Sciences* 115.27, pp. 7153–7158. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1722330115. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1722330115 (visited on 10/15/2018).

Nie, Zuqin et al. (2012). "C-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells". In: *Cell* 151.1. 00397, pp. 68–79. ISSN: 00928674. DOI: 10.1016/j.cell.2012.08.033. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867412011014 (visited on 08/23/2016).

Nikiforova, M. N. (2000). "Proximity of Chromosomal Loci That Participate in Radiation-Induced Rearrangements in Human Cells". In: *Science* 290.5489, pp. 138–141. ISSN: 00368075, 10959203. DOI: 10.1126/science.290.5489.138. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.290.5489.138 (visited on 04/06/2019).

Nikolov, D. B. and S. K. Burley (1997). "RNA Polymerase II Transcription Initiation: A Structural View". In: *Proceedings of the National Academy of Sciences* 94.1, pp. 15–22. ISSN: 0027-8424, 1091-6490. pmid: 8990153. URL: http://www.pnas.org/content/94/1/15 (visited on 09/28/2017).

Nir, Guy et al. (2018). "Walking along Chromosomes with Super-Resolution Imaging, Contact Maps, and Integrative Modeling". In: *BioRxiv*. DOI: 10.1101/374058. URL: http://biorxiv.org/lookup/doi/10.1101/374058 (visited on 08/23/2018).

Niwa, H. (2007). "Open Conformation Chromatin and Pluripotency". In: *Genes & Development* 21.21. 00000, pp. 2671–2676. ISSN: 0890-9369. DOI: 10.1101/gad.1615707. URL: http://www.genesdev.org/cgi/doi/10.1101/gad.1615707 (visited on 08/23/2016).

Nora, Elphège P., Anton Goloborodko, et al. (2017). "Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization". In: *Cell* 169.5. 00001, 930–944.e22. ISSN: 00928674. DOI: 10.1016/j.cell.2017.05.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867417305317 (visited on 06/03/2017).

Nora, Elphège P., Bryan R. Lajoie, et al. (2012). "Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre". In: *Nature* 485.7398, pp. 381–385. ISSN: 1476-4687. DOI: 10.1038/nature11049. URL: https://www.nature.com/articles/nature11049 (visited on 04/06/2019).

Normanno, Davide et al. (2015). "Probing the Target Search of DNA-Binding Proteins in Mammalian Cells Using TetR as Model Searcher". In: *Nature Communications* 6. 00020, p. 7357. ISSN: 2041-1723. DOI: 10.1038/ncomms8357. URL: http://www.nature.com/doifinder/10.1038/ncomms8357 (visited on 02/08/2017).

Nott, Timothy J. et al. (2015). "Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles". In: *Molecular Cell* 57.5, pp. 936–947. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2015.01.013. pmid: 25747659,25747659. URL: http://www.cell.com/molecular-cell/abstract/S1097-2765(15)00014-3 (visited on 11/18/2017).

Nozaki, Tadasu et al. (2017). "Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging". In: *Molecular Cell* 67.2. 00000, 282–293.e7. ISSN: 10972765. DOI: 10.1016/j.molcel.2017.06.018. URL: http://linkinghub.elsevier.com/retrieve/pii/S1097276517304458 (visited on 07/24/2017).

Nuebler, Johannes, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid Mirny (2017). "Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation". In: *bioRxiv*, p. 196261. DOI: 10.1101/196261. URL: https://www.biorxiv.org/content/early/2017/10/03/196261 (visited on 05/18/2018).

Nuebler, Johannes, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A. Mirny (2018). "Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation". In: *Proceedings of the National Academy of Sciences* 115.29, E6697–E6706. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1717730115. pmid: 29967174. URL: http://www.pnas.org/content/115/29/E6697 (visited on 11/16/2018).

Ogata, Yosihiko and Koiti Katsura (1991). "Maximum Likelihood Estimates of the Fractal Dimension for Random Spatial Patterns". In: *Biometrika*, pp. 463–474.

Oomen, Marlies E. et al. (2019). "CTCF Sites Display Cell Cycle–Dependent Dynamics in Factor Binding and Nucleosome Positioning". In: *Genome Research* 29.2, pp. 236–249. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.241547.118. pmid: 30655336. URL: http://genome.cshlp.org/content/29/2/236 (visited on 04/06/2019).

Osborne, Cameron S et al. (2004). "Active Genes Dynamically Colocalize to Shared Sites of Ongoing Transcription". In: *Nature Genetics* 36.10. 00850, pp. 1065–1071. ISSN: 1061-4036. DOI: 10.1038/ng1423. URL: http://www.nature.com/doifinder/10.1038/ng1423 (visited on 12/15/2016).

O'Shaughnessy, Ben and Itamar Procaccia (1985). "Analytical Solutions for Diffusion on Fractal Objects". In: *Physical Review Letters* 54.5, pp. 455–458. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.54.455. URL: https://link.aps.org/doi/10.1103/PhysRevLett.54.455 (visited on 04/27/2019).

Osmanović, Dino and Yitzhak Rabin (2016). "Effect of Non-Specific Interactions on Formation and Stability of Specific Complexes". In: *The Journal of Chemical Physics* 144.20. 00000, p. 205104. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.4952981. URL: http://scitation.aip.org/content/aip/journal/jcp/144/20/10.1063/1.4952981 (visited on 07/13/2016).

Ou, Horng D. et al. (2017). "ChromEMT: Visualizing 3D Chromatin Structure and Compaction in Interphase and Mitotic Cells". In: *Science* 357.6349, eaag0025. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aag0025. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aag0025 (visited on 09/20/2017).

Ovesnỳ, Martin (2016). "Vỳpocetní Metody v Jednomolekulové Lokalizacní Mikroskopii Computational Methods in Single Molecule Localization Microscopy". 00000. URL: https://www.researchgate.net/profile/Martin_Ovesny/publication/311426573_Computational_methods_in_single_molecule_localization_microscopy/links/5845ad5c08ae2d2175681aeb.pdf (visited on 07/21/2017).

Padinhateeri, Ranjith and John F. Marko (2011). "Nucleosome Positioning in a Model of Active Chromatin Remodeling Enzymes". In: *Proceedings of the National Academy of Sciences* 108.19, pp. 7799–7803. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1015206108. pmid: 21518900. URL: https://www.pnas.org/content/108/19/7799 (visited on 04/06/2019).

Paez-Segala, Maria G. et al. (2015). "Fixation-Resistant Photoactivatable Fluorescent Proteins for CLEM". In: *Nature Methods* 12.3, pp. 215–218. ISSN: 1548-7105. DOI: 10.1038/nmeth.3225. URL: https://www.nature.com/articles/nmeth.3225 (visited on 07/17/2018).

Pageon, Sophie V. et al. (2016). "Clus-DoC: A Combined Cluster Detection and Colocalization Analysis for Single-Molecule Localization Microscopy Data". In: *Molecular Biology of the Cell* 27.22. 00000, pp. 3627–3636. URL: http://www.molbiolcell.org/content/27/22/3627.short (visited on 12/02/2016).

Pak, Chi W. et al. (2016). "Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein". In: *Molecular Cell* 63.1, pp. 72–85. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2016.05.042. URL: http://www.cell.com/molecular-cell/abstract/S1097-2765(16)30228-3 (visited on 11/05/2017).

Palancade, Benoît and Olivier Bensaude (2003). "Investigating RNA Polymerase II Carboxyl-Terminal Domain (CTD) Phosphorylation". In: *European Journal of Biochemistry* 270.19, pp. 3859–3870. ISSN: 1432-1033. DOI: 10.1046/j.1432-1033.2003.03794.x. URL: http://onlinelibrary.wiley.com/doi/10.1046/j.1432-1033.2003.03794.x/abstract (visited on 11/21/2017).

Pardee, Arthur, François Jacob, and Jacques Monod (1959). "The Genetic Control and Cytoplasmic Expression of" Inducibility" in the Synthesis Of∼. Galactosidase by E. CoUt". In: *J. Mol. Biol* 1, pp. 165–178. URL: http://www.sciencedirect.com/science/article/pii/S0022283659800450/pdf?md5=4ab50fc991670adb2a39fd07aa6228fc&isDTMRedir=Y&pid=1-s2.0-S0022283659800450-main.pdf&_valck=1 (visited on 09/27/2017).

Parmar, Jyotsana J., John F. Marko, and Ranjith Padinhateeri (2014). "Nucleosome Positioning and Kinetics near Transcription-Start-Site Barriers Are Controlled by Interplay between Active Remodeling and DNA Sequence". In: *Nucleic Acids Research* 42.1, pp. 128–136. ISSN: 0305-1048. DOI: 10.1093/nar/gkt854. URL: https://academic.oup.com/nar/article/42/1/128/2436072 (visited on 04/06/2019).

Parmar, Jyotsana J., Maxime Woringer, and Christophe Zimmer (2019). "How the Genome Folds: The Biophysics of Four-Dimensional Chromatin Organization". In: *Annual Review of Biophysics* 48.1, null. DOI: 10.1146/annurev-biophys-052118-115638. pmid: 30835504. URL: https://doi.org/10.1146/annurev-biophys-052118-115638 (visited on 03/12/2019).

Patel, Avinash, Hyun O. Lee, et al. (2015). "A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation". In: *Cell* 162.5. 00146, pp. 1066–1077. ISSN: 00928674. DOI: 10.1016/j.cell.2015.07.047. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867415009630 (visited on 05/19/2017).

Patel, Avinash, Liliana Malinovska, et al. (2017). "ATP as a Biological Hydrotrope". In: *Science* 356.6339, pp. 753–756. URL: http://science.sciencemag.org/content/356/6339/753.abstract (visited on 09/12/2017).

Pelengaris, Stella, Mike Khan, and Gerard Evan (2002). "C-MYC: More than Just a Matter of Life and Death". In: *Nature Reviews Cancer* 2.10, pp. 764–776. ISSN: 1474-1768. DOI: 10.1038/nrc904. URL: https://www.nature.com/articles/nrc904 (visited on 03/30/2019).

Peng, Qin et al. (2018). "Coordinated Histone Modifications and Chromatin Reorganization in a Single Cell Revealed by FRET Biosensors". In: *Proceedings of the National Academy of Sciences* 115.50, E11681–E11690. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1811818115. pmid: 30478057. URL: https://www.pnas.org/content/115/50/E11681 (visited on 12/12/2018).

Pereira, Maria Carolina F et al. (2018). "Complementary Chromosome Folding by Transcription Factors and Cohesin". In: *BioRxiv*. DOI: 10.1101/305359. URL: http://biorxiv.org/lookup/doi/10.1101/305359 (visited on 04/23/2018).

Pérez-Cruz, Fernando (2008). "Kullback-Leibler Divergence Estimation of Continuous Distributions". In: *Information Theory, 2008. ISIT 2008. IEEE International Symposium On*. IEEE, pp. 1666–1670.

Periasamy, N and A S Verkman (1998). "Analysis of Fluorophore Diffusion by Continuous Distributions of Diffusion Coefficients: Application to Photobleaching Measurements of Multicomponent and Anomalous Diffusion". In: *Biophysical Journal* 75, p. 11.

Persson, Fredrik et al. (2013). "Extracting Intracellular Diffusive States and Transition Rates from Single-Molecule Tracking Data". In: *Nature Methods* 10.3. 00085, pp. 265–269. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.2367. URL: http://www.nature.com/doifinder/10.1038/nmeth.2367 (visited on 11/17/2016).

Phair, Robert D., Stanislaw A. Gorski, and Tom Misteli (2003). "Measurement of Dynamic Protein Binding to Chromatin In Vivo, Using Photobleaching Microscopy". In: *Methods in Enzymology*. Vol. 375. Supplement C vols. Chromatin and Chromatin Remodeling Enzymes, Part A. Academic Press, pp. 393–414. DOI: 10.1016/S0076-6879(03)75025-3. URL: http://www.sciencedirect.com/science/article/pii/S0076687903750253 (visited on 10/17/2017).

Phair, Robert D. and Tom Misteli (2001). "Kinetic Modelling Approaches to in Vivo Imaging". In: *Nature reviews. Molecular cell biology* 2.12. 00278, p. 898. URL: http://search.proquest.com/openview/546333a49d1c47ef01758fa0246cb2ad/1?pq-origsite=gscholar&cbl=27585 (visited on 07/15/2017).

Pike, Jeremy A et al. (2018). "Topological Data Analysis Quantifies Biological Nano-Structure from Single Molecule Localization Microscopy". In: *BioRxiv*. DOI: 10.1101/400275. URL: http://biorxiv.org/lookup/doi/10.1101/400275 (visited on 09/15/2018).

Poorey, K. et al. (2013). "Measuring Chromatin Interaction Dynamics on the Second Time Scale at Single-Copy Genes". In: *Science* 342.6156, pp. 369–372. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1242369. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1242369 (visited on 08/23/2016).

Pope, Benjamin D. et al. (2014). "Topologically Associating Domains Are Stable Units of Replication-Timing Regulation". In: *Nature* 515.7527, pp. 402–405. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature13986. URL: http://www.nature.com/articles/nature13986 (visited on 04/06/2019).

Pradeepa, Madapura M et al. (2016). "Histone H3 Globular Domain Acetylation Identifies a New Class of Enhancers". In: *Nature Genetics* 48.6, pp. 681–686. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/ng.3550. URL: http://www.nature.com/articles/ng.3550 (visited on 03/04/2018).

Presman, Diego M. et al. (2017). "Quantifying Transcription Factor Binding Dynamics at the Single-Molecule Level in Live Cells". In: *Methods*. The 4D Nucleome 123, pp. 76–88. ISSN: 1046-2023. DOI: 10.1016/j.ymeth.2017.03.014. URL: http://www.sciencedirect.com/science/article/pii/S1046202316304509 (visited on 10/31/2018).

Price, David H. (2018). "Transient Pausing by RNA Polymerase II". In: *Proceedings of the National Academy of Sciences*, p. 201805129. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1805129115. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1805129115 (visited on 05/17/2018).

Przybylski, Adrian et al. (2017). "Gpufit: An Open-Source Toolkit for GPU-Accelerated Curve Fitting". In: *bioRxiv*. 00000. DOI: 10.1101/174110. URL: http://biorxiv.org/content/early/2017/08/10/174110.abstract.

Pueschel, R., F. Coraggio, and P. Meister (2016). "From Single Genes to Entire Genomes: The Search for a Function of Nuclear Organization". In: *Development* 143.6. 00007, pp. 910–923. ISSN: 0950-1991, 1477-9129. DOI: 10.1242/dev.129007. URL: http://dev.biologists.org/cgi/doi/10.1242/dev.129007 (visited on 01/25/2017).

Qian, Hong (2000). "Single-Particle Tracking: Brownian Dynamics of Viscoelastic Materials". In: *Biophysical journal* 79.1, pp. 137–143.

Qin, Peiwu et al. (2017). "Live Cell Imaging of Low- and Non-Repetitive Chromosome Loci Using CRISPR-Cas9". In: *Nature Communications* 8. 00004, p. 14725. ISSN: 2041-1723. DOI: 10.1038/ncomms14725. URL: http://www.nature.com/doifinder/10.1038/ncomms14725 (visited on 07/15/2017).

Qu, Lei et al. (2011). "Simultaneous Recognition and Segmentation of Cells: Application in C.Elegans". In: *Bioinformatics* 27.20, pp. 2895–2902. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr480. URL: https://academic.oup.com/bioinformatics/article/27/20/2895/202112 (visited on 11/18/2017).

Quinodoz, Sofia A. et al. (2018). "Higher-Order Inter-Chromosomal Hubs Shape 3D Genome Organization in the Nucleus". In: *Cell*. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2018.05.024. URL: https://www.cell.com/cell/abstract/S0092-8674(18)30636-6 (visited on 06/10/2018).

Quiroz, Felipe García and Ashutosh Chilkoti (2015). "Sequence Heuristics to Encode Phase Behaviour in Intrinsically Disordered Protein Polymers". In: *Nature Materials* 14.11, pp. 1164–1171. ISSN: 1476-1122, 1476-4660. DOI: 10.1038/nmat4418. URL: http://www.nature.com/doifinder/10.1038/nmat4418 (visited on 11/22/2017).

Raccaud, Mahe et al. (2018). "Mitotic Chromosome Binding Predicts Transcription Factor Properties in Interphase". In: DOI: 10.1101/404723. URL: http://biorxiv.org/lookup/doi/10.1101/404723 (visited on 09/25/2018).

Racko, Dusan et al. (2018). "Transcription-Induced Supercoiling as the Driving Force of Chromatin Loop Extrusion during Formation of TADs in Interphase Chromosomes". In: *Nucleic Acids Research* 46.4, pp. 1648–1660. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkx1123. URL: https://academic.oup.com/nar/article/46/4/1648/4621335 (visited on 04/06/2019).

Raffeiner, Philipp et al. (2014). "In Vivo Quantification and Perturbation of Myc-Max Interactions and the Impact on Oncogenic Potential". In: *Oncotarget* 5.19, pp. 8869–8878. ISSN: 1949-2553. DOI: 10.18632/oncotarget.2588. URL: http://www.oncotarget.com/index.php?journal=oncotarget&page=article&op=view&path[]=2588&path[]=4710 (visited on 07/03/2018).

Rahl, Peter B. et al. (2010). "C-Myc Regulates Transcriptional Pause Release". In: *Cell* 141.3, pp. 432–445. ISSN: 0092-8674. DOI: 10.1016/j.cell.2010.03.030. pmid: 20434984. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2864022/ (visited on 09/06/2017).

Ramani, Vijay et al. (2017). "Massively Multiplex Single-Cell Hi-C". In: *Nature Methods* 14.3. 00007, pp. 263–266. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4155. URL: http://www.nature.com/doifinder/10.1038/nmeth.4155 (visited on 04/02/2017).

Rao, Suhas S. P. et al. (2017). "Cohesin Loss Eliminates All Loop Domains". In: *Cell* 171.2, 305–320.e24. ISSN: 1097-4172. DOI: 10.1016/j.cell.2017.09.026. pmid: 28985562.

Rao, Suhas S.P. et al. (2014). "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping". In: *Cell* 159.7, pp. 1665–1680. ISSN: 00928674. DOI: 10.1016/j.cell.2014.11.021. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867414014974 (visited on 03/12/2016).

Razin, Sergey V and Alexey A Gavrilov (2014). "Chromatin without the 30-Nm Fiber". In: *Epigenetics* 9.5, pp. 653–657. ISSN: 1559-2294. DOI: 10.4161/epi.28297. pmid: 24561903. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4063823/ (visited on 06/11/2018).

Récamier, Vincent (2013). "Single Particle Imaging in the Cell Nucleus: A Quantitative Approach". Université René Descartes-Paris V. URL: https://hal.archives-ouvertes.fr/tel-00998389/ (visited on 10/01/2017).

Récamier, Vincent et al. (2014). "Single Cell Correlation Fractal Dimension of Chromatin". In: *Nucleus* 5.1, pp. 75–84. ISSN: 1949-1034. DOI: 10.4161/nucl.28227. pmid: 24637833. URL: https://doi.org/10.4161/nucl.28227 (visited on 11/22/2017).

Reddy, K. L. et al. (2008). "Transcriptional Repression Mediated by Repositioning of Genes to the Nuclear Lamina". In: *Nature* 452.7184, pp. 243–247. ISSN: 1476-4687. DOI: 10.1038/nature06727. URL: https://www.nature.com/articles/nature06727 (visited on 04/06/2019).

Reichheld, Sean E. et al. (2017). "Direct Observation of Structure and Dynamics during Phase Separation of an Elastomeric Protein". In: *Proceedings of the National Academy of Sciences* 114.22, E4408–E4415. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1701877114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1701877114 (visited on 11/22/2017).

Reisser, Matthias and J. Christof M. Gebhardt (2017). "Decreasing Nuclear Volume Concentrates DNA and Enforces Transcription Factor-Chromatin Associations during Zebrafish Genome Activation". In: arXiv: 1710.03539 [physics]. URL: http://arxiv.org/abs/1710.03539 (visited on 04/10/2018).

Rhee, Hyun-Woo et al. (2013). "Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging". In: *Science* 339.6125. 00212, p. 1328. DOI: 10.1126/science.1230593. URL: http://science.sciencemag.org/content/339/6125/1328.abstract.

Rhodes, James et al. (2017). "Scc2/Nipbl Hops between Chromosomal Cohesin Rings after Loading". In: *eLife* 6. Ed. by Andrea Musacchio, e30000. ISSN: 2050-084X. DOI: 10.7554/eLife.30000. URL: https://doi.org/10.7554/eLife.30000 (visited on 04/24/2019).

Ricci, Maria Aurelia et al. (2015). "Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo". In: *Cell* 160.6, pp. 1145–1158. ISSN: 00928674. DOI: 10.1016/j.cell.2015.01.054. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867415001324 (visited on 06/11/2018).

Richards, Mark W. et al. (2016). "Structural Basis of N-Myc Binding by Aurora-A and Its Destabilization by Kinase Inhibitors". In: *Proceedings of the National Academy of Sciences* 113.48, pp. 13726–13731. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1610626113. pmid: 27837025. URL: http://www.pnas.org/content/113/48/13726 (visited on 05/28/2018).

Rickman, C. and W. A. Bickmore (2013). "Flashing a Light on the Spatial Organization of Transcription". In: *Science* 341.6146, pp. 621–622. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1242889. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1242889 (visited on 08/23/2016).

Riedel, Clement et al. (2014). "The Heat Released during Catalytic Turnover Enhances the Diffusion of an Enzyme". In: *Nature* 517.7533. 00044, pp. 227–230. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14043. URL: http://www.nature.com/doifinder/10.1038/nature14043 (visited on 07/26/2017).

Rieger, Bernd and Sjoerd Stallinga (2014). "The Lateral and Axial Localization Uncertainty in Super-Resolution Light Microscopy". In: *ChemPhysChem* 15.4. 00030, pp. 664–670. ISSN: 14394235. DOI: 10.1002/cphc.201300711. URL: http://doi.wiley.com/10.1002/cphc.201300711 (visited on 03/16/2017).

Rigano, Alessandro, Vanni Galli, et al. (2018). "An Algorithm-Centric Monte Carlo Method to Empirically Quantify Motion Type Estimation Uncertainty in Single-Particle Tracking". In: DOI: 10.1101/379255. URL: http://biorxiv.org/lookup/doi/10.1101/379255 (visited on 08/22/2018).

Rigano, Alessandro and Caterina Strambio De Castillia (2017). "Proposal for Minimum Information Guidelines to Report and Reproduce Results of Particle Tracking and Motion Analysis". In: *bioRxiv*. 00000. DOI: 10.1101/155036. URL: http://biorxiv.org/content/early/2017/07/13/155036.abstract.

Righini, Maurizio et al. (2018). "Full Molecular Trajectories of RNA Polymerase at Single Base-Pair Resolution". In: *Proceedings of the National Academy of Sciences* 115.6, pp. 1286–1291. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1719906115. pmid: 29351994. URL: http://www.pnas.org/content/115/6/1286 (visited on 02/09/2018).

Rippe, Karsten (2007). "Dynamic Organization of the Cell Nucleus". In: *Current Opinion in Genetics & Development* 17.5, pp. 373–380. ISSN: 0959437X. DOI: 10.1016/j.gde.2007.08.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0959437X0700158X (visited on 04/07/2019).

Robinson, Philip J. J. et al. (2006). "EM Measurements Define the Dimensions of the "30-Nm" Chromatin Fiber: Evidence for a Compact, Interdigitated Structure". In: *Proceedings of the National Academy of Sciences* 103.17, pp. 6506–6511. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0601212103. pmid: 16617109. URL: http://www.pnas.org/content/103/17/6506 (visited on 06/11/2018).

Robson, A., K. Burrage, and M. C. Leake (2012). "Inferring Diffusion in Single Live Cells at the Single-Molecule Level". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1611. 00037, pp. 20120029–20120029. ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2012.0029. URL: http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2012.0029 (visited on 12/06/2016).

Robustelli, Paul, Stefano Piana, and David E. Shaw (2018). "Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States". In: *Proceedings of the National Academy of Sciences* 115.21, E4758–E4766. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1800690115. pmid: 29735687. URL: http://www.pnas.org/content/115/21/E4758 (visited on 05/23/2018).

Rollins, Geoffrey C. et al. (2015). "Stochastic Approach to the Molecular Counting Problem in Superresolution Microscopy". In: *Proceedings of the National Academy of Sciences* 112.2, E110–E118. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1408071112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1408071112 (visited on 03/04/2018).

Rosa, Angelo and Ralf Everaers (2008). "Structure and Dynamics of Interphase Chromosomes". In: *PLOS Computational Biology* 4.8, e1000153. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000153. URL: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000153 (visited on 06/11/2018).

Rosa, Angelo and Christophe Zimmer (2014). "Chapter Nine - Computational Models of Large-Scale Genome Architecture". In: *International Review of Cell and Molecular Biology*. Ed. by Ronald Hancock and Kwang W. Jeon. Vol. 307. New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals. Academic Press, pp. 275–349. DOI: 10.1016/B978-0-12-800046-5.00009-6. URL: http://www.sciencedirect.com/science/article/pii/B9780128000465000096 (visited on 06/06/2018).

Rubinstein, Michael and Ralph H. Colby (2003). *Polymer Physics*. Oxford ; New York: Oxford University Press. 440 pp. ISBN: 978-0-19-852059-7.

Rudnick, Joseph and George Gaspari (1987). "The Shapes of Random Walks". In: *Science* 237.4813, pp. 384–389. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.237.4813.384. pmid: 17794340. URL: http://science.sciencemag.org/content/237/4813/384 (visited on 04/09/2018).

Rudolph, Claudia, Gerold Adam, and Andreas Simm (1999). "Determination of Copy Number of C-Myc Protein per Cell by Quantitative Western Blotting". In: *Analytical Biochemistry* 269.1. 00044, pp. 66–71. ISSN: 0003-2697. DOI: 10.1006/abio.1999.3095. URL: http://www.sciencedirect.com/science/article/pii/S0003269799930956.

Ruiz García, Yara et al. (2017). "Specific dsDNA Recognition by a Mimic of the DNA Binding Domain of the C-Myc/Max Transcription Factor". In: *Chemical Communications* 53.49, pp. 6653–6656. ISSN: 1359-7345, 1364-548X. DOI: 10.1039/C7CC01705G. URL: http://xlink.rsc.org/?DOI=C7CC01705G (visited on 05/13/2018).

Rust, Michael J, Mark Bates, and Xiaowei Zhuang (2006). "Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM)". In: *Nature Methods* 3.10, pp. 793–796. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth929. URL: http://www.nature.com/doifinder/10.1038/nmeth929 (visited on 11/18/2017).

Sabari, Benjamin R. et al. (2018). "Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control". In: *Science* 361.6400, eaar3958. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aar3958. pmid: 29930091. URL: http://science.sciencemag.org/content/361/6400/eaar3958 (visited on 09/18/2018).

Sabò, Arianna and Bruno Amati (2018). "BRD4 and MYC—Clarifying Regulatory Specificity". In: *Science* 360.6390, pp. 713–714. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aat6664. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aat6664 (visited on 08/28/2018).

Sabò, Arianna, Theresia R. Kress, et al. (2014). "Selective Transcriptional Regulation by Myc in Cellular Growth Control and Lymphomagenesis". In: *Nature* 511.7510. 00129, pp. 488–492. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature13537. URL: http://www.nature.com/doifinder/10.1038/nature13537 (visited on 04/03/2017).

Saey, Tina Hesman (2018). *It May Take a Village (of Proteins) to Turn on Genes*. URL: https://www.sciencenews.org/article/it-may-take-village-proteins-turn-genes (visited on 06/25/2018).

Sage, Daniel, Hagai Kirshner, et al. (2015). "Quantitative Evaluation of Software Packages for Single-Molecule Localization Microscopy". In: *Nature Methods* 12.8, pp. 717–724. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.3442. URL: http://www.nature.com/doifinder/10.1038/nmeth.3442 (visited on 09/20/2017).

Sage, Daniel, Thanh-An Pham, et al. (2018). "Super-Resolution Fight Club: A Broad Assessment of 2D & 3D Single-Molecule Localization Microscopy Software". In: *bioRxiv*, p. 362517. DOI: 10.1101/362517. URL: https://www.biorxiv.org/content/early/2018/07/04/362517 (visited on 07/05/2018).

Saglam, Ali and Lillian Chong (2018). "Protein-Protein Binding Pathways and Calculations of Rate Constants Using Fully Continuous Explicit Solvent Simulations". In: *BioRxiv*. DOI: 10.1101/453985. URL: http://biorxiv.org/lookup/doi/10.1101/453985 (visited on 12/07/2018).

Sainsbury, Sarah, Carrie Bernecky, and Patrick Cramer (2015). "Structural Basis of Transcription Initiation by RNA Polymerase II". In: *Nature Reviews Molecular Cell Biology* 16.3, pp. 129–143. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/nrm3952. URL: http://www.nature.com/doifinder/10.1038/nrm3952 (visited on 03/12/2016).

Sanborn, Adrian L. et al. (2015). "Chromatin Extrusion Explains Key Features of Loop and Domain Formation in Wild-Type and Engineered Genomes". In: *Proceedings of the National Academy of Sciences* 112.47, E6456–E6465. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1518552112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1518552112 (visited on 04/06/2019).

Sasaki, Kazuki et al. (2009). "Real-Time Imaging of Histone H4 Hyperacetylation in Living Cells". In: *Proceedings of the National Academy of Sciences* 106.38, pp. 16257–16262. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0902150106. pmid: 19805290. URL: http://www.pnas.org/content/106/38/16257 (visited on 09/07/2018).

Saw, John G. (1983). "Dependent Unit Vectors". In: *Biometrika* 70.3, pp. 665–671.

Saxton, Michael J. (2007a). "A Biological Interpretation of Transient Anomalous Subdiffusion. I. Qualitative Model". In: *Biophysical Journal* 92.4, pp. 1178–1191. ISSN: 00063495. DOI: 10.1529/biophysj.106.092619. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349507709293 (visited on 12/05/2017).

Saxton, Michael J. (2007b). "Modeling 2D and 3D Diffusion". In: *Methods in Membrane Lipids*. Ed. by Alex M. Dopico. Red. by John M. Walker. Vol. 400. Totowa, NJ: Humana Press, pp. 295–321. ISBN: 978-1-58829-662-7 978-1-59745-519-0. DOI: 10.1007/978-1-59745-519-0_20. URL: http://link.springer.com/10.1007/978-1-59745-519-0_20 (visited on 12/20/2017).

— (2008). "Single-Particle Tracking: Connecting the Dots". In: *Nature Methods* 5.8. 00086, pp. 671–672.

— (2012). "Wanted: A Positive Control for Anomalous Subdiffusion". In: *Biophysical Journal* 103.12, pp. 2411–2422. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2012.10.038. URL: http://www.sciencedirect.com/science/article/pii/S0006349512011927 (visited on 12/11/2018).

Sazer, Shelley and Helmut Schiessel (2018). "The Biology and Polymer Physics Underlying Large-Scale Chromosome Organization". In: *Traffic* 19.2, pp. 87–104. ISSN: 13989219. DOI: 10.1111/tra.12539. URL: http://doi.wiley.com/10.1111/tra.12539 (visited on 04/06/2019).

Sbalzarini, I.F. and P. Koumoutsakos (2005). "Feature Point Tracking and Trajectory Analysis for Video Imaging in Cell Biology". In: *Journal of Structural Biology* 151.2, pp. 182–195. ISSN: 10478477. DOI: 10.1016/j.jsb.2005.06.002. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047847705001267 (visited on 11/18/2017).

Schalbetter, Stephanie Andrea et al. (2017). "SMC Complexes Differentially Compact Mitotic Chromosomes According to Genomic Context". In: *Nature Cell Biology* 19.9, pp. 1071–1080. ISSN: 1465-7392, 1476-4679. DOI: 10.1038/ncb3594. URL: http://www.nature.com/articles/ncb3594 (visited on 04/06/2019).

Schalch, Thomas et al. (2005). "X-Ray Structure of a Tetranucleosome and Its Implications for the Chromatin Fibre". In: *Nature* 436.7047, pp. 138–141. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature03686. URL: http://www.nature.com/articles/nature03686 (visited on 04/06/2019).

Schindelin, Johannes et al. (2012). "Fiji: An Open-Source Platform for Biological-Image Analysis". In: *Nature Methods* 9.7, pp. 676–682. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.2019. URL: http://www.nature.com/doifinder/10.1038/nmeth.2019 (visited on 11/18/2017).

Schmidt, Jens C., Arthur J. Zaug, and Thomas R. Cech (2016). "Live Cell Imaging Reveals the Dynamics of Telomerase Recruitment to Telomeres". In: *Cell* 166.5, 1188–1197.e9. ISSN: 0092-8674. DOI: 10.1016/j.cell.2016.07.033. URL: http://www.sciencedirect.com/science/article/pii/S0092867416309837 (visited on 11/13/2018).

Schneider, Falk et al. (2018). "Nanoscale Spatio-Temporal Diffusion Modes Measured by Simultaneous Confocal and STED Imaging". In: *bioRxiv*. DOI: 10.1101/287680. URL: http://biorxiv.org/lookup/doi/10.1101/287680 (visited on 04/17/2018).

Schnell, Ulrike et al. (2012). "Immunolabeling Artifacts and the Need for Live-Cell Imaging". In: *Nature Methods* 9.2, pp. 152–158. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1855. URL: http://www.nature.com/doifinder/10.1038/nmeth.1855 (visited on 09/25/2017).

Schnitzbauer, Joerg, Maximilian T. Strauss, et al. (2017). "Super-Resolution Microscopy with DNA-PAINT". In: *Nature Protocols* 12.6, pp. 1198–1228. ISSN: 1750-2799. DOI: 10.1038/nprot.2017.024. URL: https://www.nature.com/articles/nprot.2017.024 (visited on 04/15/2019).

Schnitzbauer, Joerg, Yina Wang, et al. (2018). "Correlation Analysis Framework for Localization-Based Superresolution Microscopy". In: *Proceedings of the National Academy of Sciences* 115.13, pp. 3219–3224. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1711314115. pmid: 29531072. URL: http://www.pnas.org/content/115/13/3219 (visited on 03/28/2018).

Schoenfelder, Stefan et al. (2010). "Preferential Associations between Co-Regulated Genes Reveal a Transcriptional Interactome in Erythroid Cells". In: *Nature Genetics* 42.1. 00479, pp. 53–61. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/ng.496. URL: http://www.nature.com/doifinder/10.1038/ng.496 (visited on 12/15/2016).

Schorb, Martin and John A.G. Briggs (2014). "Correlated Cryo-Fluorescence and Cryo-Electron Microscopy with High Spatial Precision and Improved Sensitivity". In: *Ultramicroscopy* 143, pp. 24–32. ISSN: 03043991. DOI: 10.1016/j.ultramic.2013.10.015. URL: http://linkinghub.elsevier.com/retrieve/pii/S030439911300291X (visited on 08/26/2018).

Schüller, Roland and Dirk Eick (2016). "Getting Access to Low-Complexity Domain Modifications". In: *Trends in biochemical sciences* 41.11, pp. 894–897.

Schwarzer, Wibke et al. (2016). "Two Independent Modes of Chromosome Organization Are Revealed by Cohesin Removal". In: *bioRxiv*. 00000, p. 094185. URL: http://biorxiv.org/content/early/2016/12/15/094185.abstract (visited on 01/16/2017).

Sciutto, S. J. (1994). "Study of the Shape of Random Walks". In: *Journal of Physics A: Mathematical and General* 27.21, p. 7015. ISSN: 0305-4470. DOI: 10.1088/0305-4470/27/21/017. URL: http://stacks.iop.org/0305-4470/27/i=21/a=017 (visited on 04/09/2018).

Sears, Rosalie C. (2004). "The Life Cycle of C-Myc: From Synthesis to Degradation". In: *Cell Cycle* 3.9, pp. 1131–1135. ISSN: 1538-4101, 1551-4005. DOI: 10.4161/cc.3.9.1145. URL: http://www.tandfonline.com/doi/abs/10.4161/cc.3.9.1145 (visited on 11/09/2017).

Seitz, Volkhard et al. (2011). "Deep Sequencing of MYC DNA-Binding Sites in Burkitt Lymphoma". In: *PLoS ONE* 6.11. Ed. by Luwen Zhang, e26837. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0026837. URL: http://dx.plos.org/10.1371/journal.pone.0026837 (visited on 08/23/2016).

Sekiya, T. et al. (2009). "Nucleosome-Binding Affinity as a Primary Determinant of the Nuclear Mobility of the Pioneer Transcription Factor FoxA". In: *Genes & Development* 23.7. 00114, pp. 804–809. ISSN: 0890-9369. DOI: 10.1101/gad.1775509. URL: http://genesdev.cshlp.org/cgi/doi/10.1101/gad.1775509 (visited on 04/20/2017).

Seksek, Olivier, Joachim Biwersi, and A. S. Verkman (1997). "Translational Diffusion of Macromolecule-Sized Solutes in Cytoplasm and Nucleus". In: *The Journal of cell biology* 138.1, pp. 131–142. URL: http://jcb.rupress.org/content/138/1/131.abstract (visited on 09/22/2017).

Selenko, Philipp and Gerhard Wagner (2007). "Looking into Live Cells with In-Cell NMR Spectroscopy". In: *Journal of Structural Biology*. Structural Analysis of Supramolecular Assemblies by Hybrid Methods 158.2, pp. 244–253. ISSN: 1047-8477. DOI: 10.1016/j.jsb.2007.04.001. URL: http://www.sciencedirect.com/science/article/pii/S1047847707000792 (visited on 11/01/2017).

Sengupta, Prabuddha, Tijana Jovanovic-Talisman, and Jennifer Lippincott-Schwartz (2013). "Quantifying Spatial Organization in Point-Localization Superresolution Images Using Pair Correlation Analysis". In: *Nature Protocols* 8.2. 00057, pp. 345–354. ISSN: 1754-2189, 1750-2799. DOI: 10.1038/nprot.2013.005. URL: http://www.nature.com/doifinder/10.1038/nprot.2013.005 (visited on 07/20/2017).

Sengupta, Prabuddha, Tijana Jovanovic-Talisman, Dunja Skoko, et al. (2011). "Probing Protein Heterogeneity in the Plasma Membrane Using PALM and Pair Correlation Analysis". In: *Nature Methods* 8.11. 00242, pp. 969–975. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1704. URL: http://www.nature.com/doifinder/10.1038/nmeth.1704 (visited on 03/16/2017).

Sengupta, Samudra et al. (2013). "Enzyme Molecules as Nanomotors". In: *Journal of the American Chemical Society* 135.4, pp. 1406–1414. ISSN: 0002-7863. DOI: 10.1021/ja3091615. URL: http://dx.doi.org/10.1021/ja3091615 (visited on 10/29/2017).

Serber, Zach and Volker Dötsch (2001). "In-Cell NMR Spectroscopy". In: *Biochemistry* 40.48, pp. 14317–14323. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi011751w. URL: http://pubs.acs.org/doi/abs/10.1021/bi011751w (visited on 12/05/2017).

Sergé, Arnauld et al. (2008). "Dynamic Multiple-Target Tracing to Probe Spatiotemporal Cartography of Cell Membranes". In: *Nature Methods* 5.8, pp. 687–694. ISSN: 1548-7091. DOI: 10.1038/nmeth.1233. URL: http://www.nature.com/nmeth/journal/v5/n8/full/nmeth.1233.html?foxtrotcallback=true (visited on 09/04/2017).

Shagolsem, Lenin S. and Yitzhak Rabin (2016). "Particle Dynamics in Fluids with Random Interactions". In: *The Journal of Chemical Physics* 144.19. 00000, p. 194504. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.4949546. URL: http://scitation.aip.org/content/aip/journal/jcp/144/19/10.1063/1.4949546 (visited on 07/13/2016).

Shah, Naman B. et al. (2013). "F $_1$ -ATPase of *Escherichia Coli*: THE ϵ-INHIBITED STATE FORMS AFTER ATP HYDROLYSIS, IS DISTINCT FROM THE ADP-INHIBITED STATE, AND RESPONDS DYNAMICALLY TO CATALYTIC SITE LIGANDS". In: *Journal of Biological Chemistry* 288.13, pp. 9383–9395. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M113.451583. URL: http://www.jbc.org/lookup/doi/10.1074/jbc.M113.451583 (visited on 04/25/2019).

Shammas, Sarah L (2017). "Mechanistic Roles of Protein Disorder within Transcription". In: *Current Opinion in Structural Biology* 42, pp. 155–161. ISSN: 0959440X. DOI: 10.1016/j.sbi.2017.02.003. URL: http://linkinghub.elsevier.com/retrieve/pii/S0959440X1730026X (visited on 11/22/2017).

Shao, Shipeng, Boxin Xue, and Yujie Sun (2018). "Intranucleus Single-Molecule Imaging in Living Cells". In: *Biophysical Journal* 115.2, pp. 181–189. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2018.05.017. pmid: 29861035. URL: https://www.cell.com/biophysj/abstract/S0006-3495(18)30619-2 (visited on 07/18/2018).

Shao, Wanqing and Julia Zeitlinger (2017). "Paused RNA Polymerase II Inhibits New Transcriptional Initiation". In: *Nature Genetics*. 00000. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/ng.3867. URL: http://www.nature.com/doifinder/10.1038/ng.3867 (visited on 06/01/2017).

Sharp, Kim et al. (1987). "Brownian Dynamics Simulation of Diffusion to Irregular Bodies". In: *Journal of Physical Chemistry* 91.13. 00041, pp. 3624–3631. URL: http://pubs.acs.org/doi/pdf/10.1021/j100297a032 (visited on 06/19/2017).

Shav-Tal, Yaron et al. (2005). "Dynamic Sorting of Nuclear Components into Distinct Nucleolar Caps during Transcriptional Inhibition". In: *Molecular Biology of the Cell* 16.5, pp. 2395–2413. ISSN: 1059-1524, 1939-4586. DOI: 10.1091/mbc.E04-11-0992. pmid: 15758027. URL: http://www.molbiolcell.org/content/16/5/2395 (visited on 10/30/2017).

Shen, Hao et al. (2017). "Single Particle Tracking: From Theory to Biophysical Applications". In: *Chemical Reviews* 117.11, pp. 7331–7376. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.6b00815. URL: http://pubs.acs.org/doi/10.1021/acs.chemrev.6b00815 (visited on 04/24/2019).

Sherry, Kathryn P. et al. (2017). "Control of Transcriptional Activity by Design of Charge Patterning in the Intrinsically Disordered RAM Region of the Notch Receptor". In: *Proceedings of the National Academy of Sciences* 114.44, E9243–E9252. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1706083114. pmid: 29078291. URL: http://www.pnas.org/content/114/44/E9243 (visited on 10/31/2017).

Shevchenko, Georgiy (2015). "Fractional Brownian Motion in a Nutshell". In: *International Journal of Modern Physics: Conference Series*. Vol. 36. World Scientific, p. 1560002.

Shinkai, Soya et al. (2016). "Dynamic Nucleosome Movement Provides Structural Information of Topological Chromatin Domains in Living Human Cells". In: *PLOS Computational Biology* 12.10. Ed. by Alexandre V Morozov. 00005,

e1005136. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005136. URL: http://dx.plos.org/10.1371/journal.pcbi.1005136 (visited on 06/26/2017).

— (2017). "Bridging the Dynamics and Organization of Chromatin Domains by Mathematical Modeling". In: *Nucleus*. 00000, pp. 1–7. ISSN: 1949-1034, 1949-1042. DOI: 10.1080/19491034.2017.1313937. URL: https://www.tandfonline.com/doi/full/10.1080/19491034.2017.1313937 (visited on 06/26/2017).

Shivanandan, Arun, Aleksandra Radenovic, and Ivo F. Sbalzarini (2013). "MosaicIA: An ImageJ/Fiji Plugin for Spatial Pattern and Interaction Analysis". In: *BMC Bioinformatics* 14.1. 00025, p. 349. ISSN: 1471-2105. DOI: 10.1186/1471-2105-14-349. URL: https://doi.org/10.1186/1471-2105-14-349.

Shkilev, Viktor P. (2018). "A Kinetic Model for Fluorescence Microscopy Experiments in Disordered Media That Contains Binding Sites and Obstacles". In: DOI: 10.1101/302422. URL: http://biorxiv.org/lookup/doi/10.1101/302422 (visited on 06/13/2018).

Shukron, Ofir and David Holcman (2017). "Two Loci Single Particle Trajectories Analysis: Constructing A First Passage Time Statistics Of Local Chromatin Exploration". In: *bioRxiv*. 00000. DOI: 10.1101/135012. URL: http://biorxiv.org/content/early/2017/05/06/135012.abstract.

Sicorello, Alessandro et al. (2018). "A Strategy to Study Intrinsically Mixed Folded Proteins: The Structure in Solution of Ataxin-3". In: *BioRxiv*, p. 25.

Siemons, Marijn et al. (2018). "High Precision Wavefront Control in Point Spread Function Engineering for Single Emitter Localization". In: *bioRxiv*, p. 267864. DOI: 10.1101/267864. URL: https://www.biorxiv.org/content/early/2018/02/19/267864 (visited on 02/20/2018).

Singh, Anand P. et al. (2017). "3D Protein Dynamics in the Cell Nucleus". In: *Biophysical Journal* 112.1. 00001, pp. 133–142. ISSN: 00063495. DOI: 10.1016/j.bpj.2016.11.3196. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349516342655 (visited on 04/02/2017).

Sinha, Deepak Kumar et al. (2008). "Probing the Dynamic Organization of Transcription Compartments and Gene Loci within the Nucleus of Living Cells". In: *Biophysical Journal* 95.11. 00014, pp. 5432–5438. ISSN: 00063495. DOI: 10.1529/biophysj.108.135921. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349508789665 (visited on 08/23/2016).

Sinha, Kalyan K., John D. Gross, and Geeta J. Narlikar (2017). "Distortion of Histone Octamer Core Promotes Nucleosome Mobilization by a Chromatin Remodeler". In: *Science* 355.6322. 00005, eaaa3761. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aaa3761. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aaa3761 (visited on 07/18/2017).

Slator, Paddy J., Christopher W. Cairo, and Nigel J. Burroughs (2015). "Detection of Diffusion Heterogeneity in Single Particle Tracking Trajectories Using a Hidden Markov Model with Measurement Noise Propagation". In: *PLOS ONE* 10.10. Ed. by Marek Cebecauer. 00002, e0140759. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0140759. URL: http://dx.plos.org/10.1371/journal.pone.0140759 (visited on 11/10/2016).

Slator, Paddy John and Nigel Burroughs (2018). "A Hidden Markov Model for Detecting Confinement in Single Particle Tracking Trajectories". In: *bioRxiv*, p. 275107. DOI: 10.1101/275107. URL: https://www.biorxiv.org/content/early/2018/03/08/275107 (visited on 03/09/2018).

Smith, Keriayn N., Amar M. Singh, and Stephen Dalton (2010). "Myc Represses Primitive Endoderm Differentiation in Pluripotent Stem Cells". In: *Cell Stem Cell* 7.3. 00144, pp. 343–354. ISSN: 19345909. DOI: 10.1016/j.stem.2010.06.023. URL: http://linkinghub.elsevier.com/retrieve/pii/S1934590910003504 (visited on 08/03/2017).

So, Min-kyung, Hequan Yao, and Jianghong Rao (2008). "HaloTag Protein-Mediated Specific Labeling of Living Cells with Quantum Dots". In: *Biochemical and Biophysical Research Communications* 374.3. 00060, pp. 419–423. ISSN: 0006291X. DOI: 10.1016/j.bbrc.2008.07.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006291X08012989 (visited on 08/23/2016).

Solovei, Irina et al. (2009). "Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution". In: *Cell* 137.2, pp. 356–368. ISSN: 00928674. DOI: 10.1016/j.cell.2009.01.052. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867409001378 (visited on 04/06/2019).

Song, F. et al. (2014). "Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units". In: *Science* 344.6182, pp. 376–380. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1251413. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1251413 (visited on 04/06/2019).

Soufi, Abdenour et al. (2015). "Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming". In: *Cell* 161.3. 00093, pp. 555–568. ISSN: 00928674. DOI: 10.1016/j.cell.2015.03.017. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867415003049 (visited on 04/20/2017).

Spahn, Philipp N et al. (2017). "PinAPL-Py: A Web-Service for the Analysis of CRISPR-Cas9 Screens". In: *bioRxiv*. 00000. DOI: 10.1101/147462. URL: http://biorxiv.org/content/early/2017/06/08/147462.abstract.

Spector, David L and Susan M Gasser (2003). "A Molecular Dissection of Nuclear Function: Conference on The Dynamic Nucleus: Questions and Implications". In: *EMBO reports* 4.1, pp. 18–23. ISSN: 1469221X. DOI: 10.1038/sj.embor.embor701. URL: http://embor.embopress.org/cgi/doi/10.1038/sj.embor.embor701 (visited on 04/06/2019).

Spector, David L. and Angus I. Lamond (2011). "Nuclear Speckles". In: *Cold Spring Harbor Perspectives in Biology* 3.2, a000646. ISSN: , 1943-0264. DOI: 10.1101/cshperspect.a000646. pmid: 20926517. URL: http://cshperspectives.cshlp.org/content/3/2/a000646 (visited on 06/12/2018).

Speil, Jasmin et al. (2011). "Activated STAT1 Transcription Factors Conduct Distinct Saltatory Movements in the Cell Nucleus". In: *Biophysical Journal* 101.11, pp. 2592–2600. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2011.10.006. URL: http://www.sciencedirect.com/science/article/pii/S0006349511011957 (visited on 02/05/2018).

Spille, Jan-Hendrik et al. (2012). "Dynamic Three-Dimensional Tracking of Single Fluorescent Nanoparticles Deep inside Living Tissue". In: *Optics Express* 20.18, pp. 19697–19707. ISSN: 1094-4087. DOI: 10.1364/OE.20.019697. URL: https://www.osapublishing.org/abstract.cfm?uri=oe-20-18-19697 (visited on 09/12/2017).

Spotts, GERALD D. and STEPHEN R. Hann (1990). "Enhanced Translation and Increased Turnover of C-Myc Proteins Occur during Differentiation of Murine Erythroleukemia Cells." In: *Molecular and cellular biology* 10.8. 00083, pp. 3952–3964. URL: http://mcb.asm.org/content/10/8/3952.short (visited on 04/20/2017).

Sprouse, Rebekka O. et al. (2008). "Regulation of TATA-Binding Protein Dynamics in Living Yeast Cells". In: *Proceedings of the National Academy of Sciences* 105.36, pp. 13304–13308. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0801901105. pmid: 18765812. URL: http://www.pnas.org/content/105/36/13304 (visited on 09/18/2018).

Stadhouders, Ralph et al. (2018). "Transcription Factors Orchestrate Dynamic Interplay between Genome Topology and Gene Regulation during Cell Reprogramming". In: *Nature Genetics* 50.2, pp. 238–249. ISSN: 1546-1718. DOI: 10.1038/s41588-017-0030-7. URL: https://www.nature.com/articles/s41588-017-0030-7 (visited on 02/12/2018).

Stasevich, Timothy J. et al. (2014). "Regulation of RNA Polymerase II Activation by Histone Acetylation in Single Living Cells". In: *Nature* 516.7530, pp. 272–275. ISSN: 1476-4687. DOI: 10.1038/nature13714. URL: https://www.nature.com/articles/nature13714 (visited on 06/12/2018).

Stein, C. A. et al. (2010). "Efficient Gene Silencing by Delivery of Locked Nucleic Acid Antisense Oligonucleotides, Unassisted by Transfection Reagents". In: *Nucleic Acids Research* 38.1, e3. ISSN: 0305-1048. DOI: 10.1093/nar/gkp841. pmid: 19854938. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2800216/ (visited on 08/24/2018).

Steurer, Barbara et al. (2018). "Live-Cell Analysis of Endogenous GFP-RPB1 Uncovers Rapid Turnover of Initiating and Promoter-Paused RNA Polymerase II". In: *Proceedings of the National Academy of Sciences* 115.19, E4368–E4376. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1717920115. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1717920115 (visited on 06/22/2018).

Stojanova, Angelina et al. (2016). "MYC Interaction with the Tumor Suppressive SWI/SNF Complex Member INI1 Regulates Transcription and Cellular Transformation". In: *Cell Cycle* 15.13. 00003, pp. 1693–1705. ISSN: 1538-4101, 1551-4005. DOI: 10.1080/15384101.2016.1146836. URL: http://www.tandfonline.com/doi/full/10.1080/15384101.2016.1146836 (visited on 04/28/2017).

Stracy, Mathew et al. (2016). "Single-Molecule Imaging of UvrA and UvrB Recruitment to DNA Lesions in Living Escherichia Coli". In: *Nature Communications* 7. 00000, p. 12568. ISSN: 2041-1723. DOI: 10.1038/ncomms12568. URL: http://www.nature.com/doifinder/10.1038/ncomms12568 (visited on 01/23/2017).

Strickland, Devin et al. (2012). "TULIPs: Tunable, Light-Controlled Interacting Protein Tags for Cell Biology". In: *Nature Methods* 9.4. 00138, pp. 379–384. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.1904. URL: http://www.nature.com/doifinder/10.1038/nmeth.1904 (visited on 08/23/2016).

Strom, Amy R. et al. (2017). "Phase Separation Drives Heterochromatin Domain Formation". In: *Nature* 547.7662. 00001, pp. 241–245. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature22989. URL: http://www.nature.com/doifinder/10.1038/nature22989 (visited on 07/15/2017).

Stützer, Alexandra et al. (2016). "Modulations of DNA Contacts by Linker Histones and Post-Translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails". In: *Molecular Cell* 61.2, pp. 247–259. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2015.12.015. URL: http://www.sciencedirect.com/science/article/pii/S1097276515009715 (visited on 10/30/2017).

Sukenik, Shahar, Pin Ren, and Martin Gruebele (2017). "Weak Protein-Protein Interactions in Live Cells Are Quantified by Cell-Volume Modulation". In: *Proceedings of the National Academy of Sciences*. 00000, p. 201700818. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1700818114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1700818114 (visited on 06/27/2017).

Sumida, Noriyuki et al. (2018). "The Ultra-Sensitive Nodewalk Technique Identifies Stochastic from Virtual, Population-Based Enhancer Hubs Regulating MYC in 3D: Implications for the Fitness of Cancer Cells". In: DOI: 10.1101/286583. URL: http://biorxiv.org/lookup/doi/10.1101/286583 (visited on 03/29/2018).

Sun, Guangyu et al. (2015). "Bayesian Model Selection Applied to the Analysis of Fluorescence Correlation Spectroscopy Data of Fluorescent Proteins *in Vitro* and *in Vivo*". In: *Analytical Chemistry* 87.8. 00000, pp. 4326–4333. ISSN: 0003-2700, 1520-6882. DOI: 10.1021/acs.analchem.5b00022. URL: http://pubs.acs.org/doi/abs/10.1021/acs.analchem.5b00022 (visited on 11/17/2016).

Sun, Ruoxi, Evan Archer, and Liam Paninski (2016). "Scalable Variational Inference for Super Resolution Microscopy". In: DOI: 10.1101/081703. URL: http://biorxiv.org/lookup/doi/10.1101/081703 (visited on 03/27/2018).

Sun, Ruoxi and Liam Paninski (2018). "Scalable Approximate Bayesian Inference for Particle Tracking Data". In: *BioRxiv*, p. 10.

Sun, Xiao-Xin et al. (2018). "SUMO Protease SENP1 deSUMOylates and Stabilizes C-Myc". In: *Proceedings of the National Academy of Sciences* 115.43, pp. 10983–10988. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1802932115. pmid: 30305424. URL: http://www.pnas.org/content/115/43/10983 (visited on 10/24/2018).

Swinstead, Erin E. et al. (2016). "Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions". In: *Cell* 165.3. 00024, pp. 593–605. ISSN: 00928674. DOI: 10.1016/j.cell.2016.02.067. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867416302574 (visited on 05/09/2017).

Szabo, Quentin et al. (2018). "TADs Are 3D Structural Units of Higher-Order Chromosome Organization in Drosophila". In: *SCIENCE ADVANCES*, p. 14.

Szczurek, Aleksander T. et al. (2014). "Single Molecule Localization Microscopy of the Distribution of Chromatin Using Hoechst and DAPI Fluorescent Probes". In: *Nucleus* 5.4, pp. 331–340. ISSN: 1949-1034. DOI: 10.4161/nucl.29564. pmid: 25482122. URL: https://doi.org/10.4161/nucl.29564 (visited on 11/22/2017).

Szilvási-Nagy, Márta (1984). "An Algorithm for Determining the Intersection of Two Simple Polyhedra". In: *Computer Graphics Forum*. Vol. 3. 00013. Wiley Online Library, pp. 219–225. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.1984.tb00071.x/abstract (visited on 08/23/2016).

Takahashi, Kazutoshi and Shinya Yamanaka (2006). "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors". In: *Cell* 126.4, pp. 663–676. ISSN: 0092-8674. DOI: 10.1016/j.cell.2006.07.024. URL: http://www.sciencedirect.com/science/article/pii/S0092867406009767 (visited on 04/02/2019).

Tanenbaum, Marvin E. et al. (2014). "A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging". In: *Cell* 159.3, pp. 635–646. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2014.09.039. pmid: 25307933. URL: https://www.cell.com/cell/abstract/S0092-8674(14)01227-6 (visited on 06/13/2018).

Tang, Jialei and Kyu Young Han (2018). "Clear, Extended Field-of-View Single-Molecule Imaging by Highly Inclined Swept Illumination". In: DOI: 10.1101/311753. URL: http://biorxiv.org/lookup/doi/10.1101/311753 (visited on 06/20/2018).

Tanimoto, Yoko et al. (2008). "Embryonic Stem Cells Derived from C57BL/6J and C57BL/6N Mice". In: *Comparative Medicine* 58.4, pp. 347–352. ISSN: 1532-0820. pmid: 18724776. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2706045/ (visited on 11/07/2018).

Tantale, Katjana et al. (2016). "A Single-Molecule View of Transcription Reveals Convoys of RNA Polymerases and Multi-Scale Bursting". In: *Nature Communications* 7. 00006, p. 12248. ISSN: 2041-1723. DOI: 10.1038/ncomms12248. URL: http://www.nature.com/doifinder/10.1038/ncomms12248 (visited on 02/08/2017).

Tedeschi, Antonio et al. (2013). "Wapl Is an Essential Regulator of Chromatin Structure and Chromosome Segregation". In: *Nature* 501.7468, pp. 564–568. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature12471. URL: http://www.nature.com/articles/nature12471 (visited on 04/06/2019).

Tejedor, Vincent et al. (2010). "Quantitative Analysis of Single Particle Trajectories: Mean Maximal Excursion Method". In: *Biophysical Journal* 98.7. 00118, pp. 1364–1372. ISSN: 00063495. DOI: 10.1016/j.bpj.2009.12.4282. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349509060974 (visited on 12/09/2016).

Terai, Takuya et al. (2011). "Rational Development of Caged-Biotin Protein-Labeling Agents and Some Applications in Live Cells". In: *Chemistry & Biology* 18.10. 00008, pp. 1261–1272. ISSN: 10745521. DOI: 10.1016/j.chembiol.2011.09.007. URL: http://linkinghub.elsevier.com/retrieve/pii/S1074552111003164 (visited on 08/23/2016).

Terrenoire, Edith et al. (2010). "Immunostaining of Modified Histones Defines High-Level Features of the Human Metaphase Epigenome". In: *Genome Biology* 11, R110. ISSN: 1474-760X. DOI: 10.1186/gb-2010-11-11-r110. URL: https://doi.org/10.1186/gb-2010-11-11-r110 (visited on 04/11/2018).

Testa, Anna et al. (2005). "Chromatin Immunoprecipitation (ChIP) on Chip Experiments Uncover a Widespread Distribution of NF-Y Binding CCAAT Sites Outside of Core Promoters". In: *Journal of Biological Chemistry* 280.14, pp. 13606–13615. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M414039200. URL: http://www.jbc.org/lookup/doi/10.1074/jbc.M414039200 (visited on 09/21/2018).

Teves, Sheila S, Luye An, Aarohi Bhargava-Shah, et al. (2018). "A Stable Mode of Bookmarking by TBP Recruits RNA Polymerase II to Mitotic Chromosomes".

Teves, Sheila S, Luye An, Anders S Hansen, et al. (2016). "A Dynamic Mode of Mitotic Bookmarking by Transcription Factors". In: *eLife* 5. Ed. by Karen Adelman. 00012, e22280. ISSN: 2050-084X. DOI: 10.7554/eLife.22280. URL: https://doi.org/10.7554/eLife.22280.

Thapa, Samudrajit et al. (2018). "Bayesian Analysis of Single-Particle Tracking Data Using the Nested-Sampling Algorithm: Maximum-Likelihood Model Selection Applied to Stochastic-Diffusivity Data". In: *Physical Chemistry Chemical Physics*. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/C8CP04043E. URL: http://xlink.rsc.org/?DOI=C8CP04043E (visited on 10/05/2018).

Theer, Patrick, Cyril Mongis, and Michael Knop (2014). "PSFj: Know Your Fluorescence Microscope". In: *Nature methods* 11.10. 00009, pp. 981–982. URL: https://www.nature.com/articles/nmeth.3102 (visited on 09/03/2017).

Theillet, Francois-Xavier et al. (2016). "Structural Disorder of Monomeric α-Synuclein Persists in Mammalian Cells". In: *Nature* 530.7588, pp. 45–50. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature16531. URL: http://www.nature.com/doifinder/10.1038/nature16531 (visited on 10/30/2017).

Therizols, Pierre et al. (2014). "Chromatin Decondensation Is Sufficient to Alter Nuclear Organization in Embryonic Stem Cells". In: *Science* 346.6214, pp. 1238–1242. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1259587. pmid: 25477464. URL: http://science.sciencemag.org/content/346/6214/1238 (visited on 05/18/2018).

Therizols, P. et al. (2010). "Chromosome Arm Length and Nuclear Constraints Determine the Dynamic Relationship of Yeast Subtelomeres". In: *Proceedings of the National Academy of Sciences* 107.5, pp. 2025–2030. ISSN: 0027-8424,

1091-6490. DOI: 10.1073/pnas.0914187107. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0914187107 (visited on 04/06/2019).

Thomas, Lance R. et al. (2015). "Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC". In: *Molecular Cell* 58.3, pp. 440–452. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2015.02.028. URL: http://www.sciencedirect.com/science/article/pii/S1097276515001422 (visited on 07/19/2018).

Thompson, Michael A. et al. (2010). "Three-Dimensional Tracking of Single mRNA Particles in Saccharomyces Cerevisiae Using a Double-Helix Point Spread Function". In: *Proceedings of the National Academy of Sciences*. 00080.

Thongwichian, Rossukon et al. (2015). "A Multiplexed NMR-Reporter Approach to Measure Cellular Kinase and Phosphatase Activities in Real-Time". In: *Journal of the American Chemical Society* 137.20, pp. 6468–6471. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.5b02987. URL: http://pubs.acs.org/doi/10.1021/jacs.5b02987 (visited on 10/30/2017).

Tichy, Elisia D. et al. (2012). "The Abundance of Rad51 Protein in Mouse Embryonic Stem Cells Is Regulated at Multiple Levels". In: *Stem cell research* 9.2, pp. 124–134. ISSN: 1873-5061. DOI: 10.1016/j.scr.2012.05.004. pmid: 22705496. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3412895/ (visited on 10/17/2017).

Tinevez, Jean-Yves et al. (2017). "TrackMate: An Open and Extensible Platform for Single-Particle Tracking". In: *Methods* 115. 00012, pp. 80–90. ISSN: 10462023. DOI: 10.1016/j.ymeth.2016.09.016. URL: http://linkinghub.elsevier.com/retrieve/pii/S1046202316303346 (visited on 05/30/2017).

Tokunaga, Makio, Naoko Imamoto, and Kumiko Sakata-Sogawa (2008). "Highly Inclined Thin Illumination Enables Clear Single-Molecule Imaging in Cells". In: *Nature methods* 5.2. 00450, pp. 159–161. URL: https://www.nature.com/nmeth/journal/v5/n2/abs/nmeth1171.html (visited on 06/30/2017).

Topaloglu, O. (2005). "Improved Methods for the Generation of Human Gene Knockout and Knockin Cell Lines". In: *Nucleic Acids Research* 33.18, e158–e158. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gni160. URL: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gni160 (visited on 08/23/2016).

Torchia, Maria Letizia Giardino and Jonathan D. Ashwell (2018). "Getting MAD at MYC". In: *Proceedings of the National Academy of Sciences* 115.40, pp. 9821–9823. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1813867115. pmid: 30201726. URL: http://www.pnas.org/content/115/40/9821 (visited on 10/03/2018).

Tsai, Albert et al. (2017). "Nuclear Microenvironments Modulate Transcription from Low-Affinity Enhancers". In: *eLife* 6, e28975. ISSN: 2050-084X. DOI: 10.7554/eLife.28975. URL: https://elifesciences.org/articles/28975 (visited on 11/22/2017).

Tsai, Min-Yeh et al. (2016). "Molecular Mechanism of Facilitated Dissociation of Fis Protein from DNA". In: *Journal of the American Chemical Society* 138.41. 00000, pp. 13497–13500. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.6b08416. URL: http://pubs.acs.org/doi/abs/10.1021/jacs.6b08416 (visited on 12/05/2016).

Tsekouras, Konstantinos et al. (2015). "Inferring Diffusion Dynamics from FCS in Heterogeneous Nuclear Environments". In: *Biophysical Journal* 109.1. 00001, pp. 7–17. ISSN: 00063495. DOI: 10.1016/j.bpj.2015.05.035. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349515005470 (visited on 04/02/2017).

Tsunoyama, Taka A. et al. (2018). "Super-Long Single-Molecule Tracking Reveals Dynamic-Anchorage-Induced Integrin Function". In: *Nature Chemical Biology*. ISSN: 1552-4450, 1552-4469. DOI: 10.1038/s41589-018-0032-5. URL: http://www.nature.com/articles/s41589-018-0032-5 (visited on 04/02/2018).

Tu, William B. et al. (2015). "Myc and Its Interactors Take Shape". In: *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1849.5. 00029, pp. 469–483. ISSN: 18749399. DOI: 10.1016/j.bbagrm.2014.06.002. URL: http://linkinghub.elsevier.com/retrieve/pii/S1874939914001540 (visited on 04/28/2017).

Tumbar, Tudorita and Andrew S. Belmont (2001). "Interphase Movements of a DNA Chromosome Region Modulated by VP16 Transcriptional Activator". In: *Nature Cell Biology* 3.2, pp. 134–139. ISSN: 1476-4679. DOI: 10.1038/35055033. URL: https://www.nature.com/articles/ncb0201_134 (visited on 05/18/2018).

Türkcan, Silvan, Antigoni Alexandrou, and Jean-Baptiste Masson (2012). "A Bayesian Inference Scheme to Extract Diffusivity and Potential Fields from Confined Single-Molecule Trajectories". In: *Biophysical Journal* 102.10, pp. 2288–2298. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.01.063. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349512004183 (visited on 11/18/2017).

Ulianov, Sergey V. et al. (2016). "Active Chromatin and Transcription Play a Key Role in Chromosome Partitioning into Topologically Associating Domains". In: *Genome Research* 26.1, pp. 70–84. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.196006.115. URL: http://genome.cshlp.org/lookup/doi/10.1101/gr.196006.115 (visited on 04/06/2019).

Urnov, Fyodor D. et al. (2010). "Genome Editing with Engineered Zinc Finger Nucleases". In: *Nature Reviews Genetics* 11.9. 01052, pp. 636–646. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg2842. URL: http://www.nature.com/doifinder/10.1038/nrg2842 (visited on 08/23/2016).

Uslu, Veli Vural et al. (2014). "Long-Range Enhancers Regulating Myc Expression Are Required for Normal Facial Morphogenesis". In: *Nature Genetics* 46.7, pp. 753–758. ISSN: 1061-4036. DOI: 10.1038/ng.2971. URL: http://www.nature.com/ng/journal/v46/n7/full/ng.2971.html (visited on 09/07/2017).

Van de Linde, Sebastian and Markus Sauer (2014). "How to Switch a Fluorophore: From Undesired Blinking to Controlled Photoswitching". In: *Chem. Soc. Rev.* 43.4. 00036, pp. 1076–1087. ISSN: 0306-0012, 1460-4744. DOI: 10.1039/C3CS60195A. URL: http://xlink.rsc.org/?DOI=C3CS60195A (visited on 08/23/2016).

Van Arensbergen, Joris, Vincent D FitzPatrick, et al. (2016). "Genome-Wide Mapping of Autonomous Promoter Activity in Human Cells". In: *Nature Biotechnology*. 00000. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.3754. URL: http://www.nature.com/doifinder/10.1038/nbt.3754 (visited on 01/30/2017).

Van Arensbergen, Joris, Bas van Steensel, and Harmen J. Bussemaker (2014). "In Search of the Determinants of Enhancer–Promoter Interaction Specificity". In: *Trends in Cell Biology* 24.11. 00040, pp. 695–702. ISSN: 09628924. DOI: 10.1016/j.tcb.2014.07.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S0962892414001184 (visited on 12/23/2016).

Van Steensel, Bas and Andrew S. Belmont (2017). "Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression". In: *Cell* 169.5, pp. 780–791. ISSN: 00928674. DOI: 10.1016/j.cell.2017.04.022. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867417304737 (visited on 06/20/2018).

Van Steensel, Bas and Steven Henikoff (2000). "Identification of in Vivo DNA Targets of Chromatin Proteins Using Tethered Dam Methyltransferase". In: *Nature biotechnology* 18.4. 00000, pp. 424–428. URL: http://www.nature.com/nbt/journal/v18/n4/full/nbt0400_424.html (visited on 08/23/2016).

Varlakhanova, Natalia V. et al. (2010). "Myc Maintains Embryonic Stem Cell Pluripotency and Self-Renewal". In: *Differentiation* 80.1. 00063, pp. 9–19. ISSN: 03014681. DOI: 10.1016/j.diff.2010.05.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S0301468110000484 (visited on 08/03/2017).

Veatch, Sarah L. et al. (2012). "Correlation Functions Quantify Super-Resolution Images and Estimate Apparent Clustering Due to Over-Counting". In: *PLoS ONE* 7.2. Ed. by Jianghong Rao. 00137, e31457. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0031457. URL: http://dx.plos.org/10.1371/journal.pone.0031457 (visited on 03/16/2017).

Vekilov, Peter G. (2010). "Phase Transitions of Folded Proteins". In: *Soft Matter* 6.21, pp. 5254–5272.

Vestergaard, Christian L., Paul C. Blainey, and Henrik Flyvbjerg (2014). "Optimal Estimation of Diffusion Coefficients from Single-Particle Trajectories". In: *Physical Review E* 89.2. 00041. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.89.022726. URL: http://link.aps.org/doi/10.1103/PhysRevE.89.022726 (visited on 11/10/2016).

Vian, Laura et al. (2018). "The Energetics and Physiological Impact of Cohesin Extrusion". In: *Cell*. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.03.072. URL: http://www.sciencedirect.com/science/article/pii/S0092867418304045 (visited on 05/10/2018).

Vitalis, Andreas and Rohit V. Pappu (2009). "ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions". In: *Journal of Computational Chemistry* 30.5, pp. 673–699. ISSN: 1096-987X. DOI: 10.1002/jcc.21005. URL: http://onlinelibrary.wiley.com/doi/10.1002/jcc.21005/abstract (visited on 11/05/2017).

Vogel, Christine and Edward M Marcotte (2008). "Calculating Absolute and Relative Protein Abundance from Mass Spectrometry-Based Protein Expression Data". In: *Nature Protocols* 3.9. 00078, pp. 1444–1451. ISSN: 1754-2189, 1750-2799. DOI: 10.1038/nprot.2008.132. URL: http://www.nature.com/doifinder/10.1038/nprot.2008.132 (visited on 05/09/2017).

Vogelsang, Jan et al. (2008). "A Reducing and Oxidizing System Minimizes Photobleaching and Blinking of Fluorescent Dyes". In: *Angewandte Chemie International Edition* 47.29, pp. 5465–5469. ISSN: 14337851, 15213773. DOI: 10.1002/anie.200801518. URL: http://doi.wiley.com/10.1002/anie.200801518 (visited on 04/25/2019).

Von Paleske, Lisa et al. (2014). *A Novel Enhancer Region 1.7 Mb Downstream of the C-Myc Gene Drives Its Expression in Hematopoietic Stem and Progenitor Cells*. Am Soc Hematology. URL: http://www.bloodjournal.org/content/124/21/766.abstract (visited on 09/07/2017).

Von Smoluchowski, M. (1906). "Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen". In: *Annalen der Physik* 326.14, pp. 756–780. ISSN: 00033804, 15213889. DOI: 10.1002/andp.19063261405. URL: http://doi.wiley.com/10.1002/andp.19063261405 (visited on 11/21/2017).

Walther, Nike et al. (2018). "A Quantitative Map of Human Condensins Provides New Insights into Mitotic Chromosome Architecture". In: *J Cell Biol*, jcb.201801048. ISSN: 0021-9525, 1540-8140. DOI: 10.1083/jcb.201801048. URL: http://jcb.rupress.org/content/early/2018/04/06/jcb.201801048 (visited on 04/09/2018).

Walz, Susanne et al. (2014). "Activation and Repression by Oncogenic MYC Shape Tumour-Specific Gene Expression Profiles". In: *Nature* 511.7510, pp. 483–487. ISSN: 1476-4687. DOI: 10.1038/nature13473. URL: https://www.nature.com/articles/nature13473 (visited on 05/28/2018).

Wang, Fangpo and Alan E. Gelfand (2013). "Directional Data Analysis under the General Projected Normal Distribution". In: *Statistical Methodology* 10.1, pp. 113–127. ISSN: 15723127. DOI: 10.1016/j.stamet.2012.07.005. URL: http://linkinghub.elsevier.com/retrieve/pii/S1572312712000457 (visited on 10/29/2017).

— (2014). "Modeling Space and Space-Time Directional Data Using Projected Gaussian Processes". In: *Journal of the American Statistical Association* 109.508, pp. 1565–1580. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2014.934454. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.2014.934454 (visited on 10/29/2017).

Wang, Jennifer T. et al. (2014). "Regulation of RNA Granule Dynamics by Phosphorylation of Serine-Rich, Intrinsically Disordered Proteins in C. Elegans". In: *eLife* 3, e04591. ISSN: 2050-084X. DOI: 10.7554/eLife.04591. URL: https://elifesciences.org/articles/04591 (visited on 02/01/2018).

Wang, Q. and W. E. Moerner (2010). "Optimal Strategy for Trapping Single Fluorescent Molecules in Solution Using the ABEL Trap". In: *Applied Physics B* 99.1-2, pp. 23–30. ISSN: 0946-2171, 1432-0649. DOI: 10.1007/s00340-009-3843-y. URL: http://link.springer.com/10.1007/s00340-009-3843-y (visited on 04/25/2019).

Wang, Quan, Randall H. Goldsmith, et al. (2012). "Probing Single Biomolecules in Solution Using the Anti-Brownian Electrokinetic (ABEL) Trap". In: *Accounts of Chemical Research* 45.11, pp. 1955–1964. ISSN: 0001-4842, 1520-4898. DOI: 10.1021/ar200304t. URL: http://pubs.acs.org/doi/10.1021/ar200304t (visited on 04/25/2019).

Wang, Quan and W. E. Moerner (2011). "An Adaptive Anti-Brownian Electrokinetic Trap with Real-Time Information on Single-Molecule Diffusivity and Mobility". In: *ACS Nano* 5.7, pp. 5792–5799. ISSN: 1936-0851, 1936-086X. DOI: 10.1021/nn2014968. URL: http://pubs.acs.org/doi/10.1021/nn2014968 (visited on 04/25/2019).

Wang, Quan and W E Moerner (2014). "Single-Molecule Motions Enable Direct Visualization of Biomolecular Interactions in Solution". In: *Nature Methods* 11.5, pp. 555–558. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.2882. URL: http://www.nature.com/articles/nmeth.2882 (visited on 04/25/2019).

Wang, Quan, Andrew J. Serban, et al. (2018). "Single-Molecule Diffusometry Reveals the Nucleotide-Dependent Oligomerization Pathways of *Nicotiana Tabacum* Rubisco Activase". In: *The Journal of Chemical Physics* 148.12, p. 123319. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.5005930. URL: http://aip.scitation.org/doi/10.1063/1.5005930 (visited on 04/25/2019).

Wang, S. et al. (2016). "Spatial Organization of Chromatin Domains and Compartments in Single Chromosomes". In: *Science*. 00000. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aaf8084. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.aaf8084 (visited on 08/01/2016).

Wang, Yina et al. (2014). "Localization Events-Based Sample Drift Correction for Localization Microscopy with Redundant Cross-Correlation Algorithm". In: *Optics Express* 22.13. 00034, p. 15982. ISSN: 1094-4087. DOI: 10.1364/OE.22.015982. URL: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-22-13-15982 (visited on 07/20/2017).

Warfield, Linda et al. (2017). "Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID". In: *Molecular Cell*. ISSN: 10972765. DOI: 10.1016/j.molcel.2017.08.014. URL: http://linkinghub.elsevier.com/retrieve/pii/S1097276517306111 (visited on 10/06/2017).

Weber, Stephanie C. et al. (2012). "Analytical Tools To Distinguish the Effects of Localization Error, Confinement, and Medium Elasticity on the Velocity Autocorrelation Function". In: *Biophysical Journal* 102.11, pp. 2443–2450. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.03.062. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349512004031 (visited on 11/18/2017).

Wehler, Pierre et al. (2016). "Optogenetic Control of Nuclear Protein Import in Living Cells Using Light-Inducible Nuclear Localization Signals (LINuS)". In: *Current Protocols in Chemical Biology* 8.2, pp. 131–145. ISSN: 2160-4762. DOI: 10.1002/cpch.4. URL: https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpch.4 (visited on 11/13/2018).

Wei, Ming-Tzo et al. (2017). "Phase Behaviour of Disordered Proteins Underlying Low Density and High Permeability of Liquid Organelles". In: *Nature Chemistry* 9.11, pp. 1118–1125. ISSN: 1755-4330. DOI: 10.1038/nchem.2803. URL: https://www.nature.com/nchem/journal/v9/n11/full/nchem.2803.html (visited on 11/05/2017).

Weigel, Aubrey V. et al. (2012). "Obstructed Diffusion Propagator Analysis for Single-Particle Tracking". In: *Physical Review E* 85.4. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.85.041924. URL: https://link.aps.org/doi/10.1103/PhysRevE.85.041924 (visited on 02/14/2018).

Weimann, Laura et al. (2013). "A Quantitative Comparison of Single-Dye Tracking Analysis Tools Using Monte Carlo Simulations". In: *PLoS ONE* 8.5. Ed. by Attila Szolnoki. 00015, e64287. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0064287. URL: http://dx.plos.org/10.1371/journal.pone.0064287 (visited on 06/18/2017).

Weingarten-Gabbay, Shira, Ronit Nir, et al. (2017). "Deciphering Transcriptional Regulation of Human Core Promoters". In: DOI: 10.1101/174904. URL: http://biorxiv.org/lookup/doi/10.1101/174904 (visited on 09/20/2018).

Weingarten-Gabbay, Shira and Eran Segal (2014). "The Grammar of Transcriptional Regulation". In: *Human Genetics* 133.6, pp. 701–711. ISSN: 0340-6717, 1432-1203. DOI: 10.1007/s00439-013-1413-1. URL: http://link.springer.com/10.1007/s00439-013-1413-1 (visited on 11/18/2017).

Weiss, Matthias et al. (2004). "Anomalous Subdiffusion Is a Measure for Cytoplasmic Crowding in Living Cells". In: *Biophysical Journal* 87.5, pp. 3518–3524. ISSN: 00063495. DOI: 10.1529/biophysj.104.044263. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349504738163 (visited on 09/22/2017).

Weissman, Haim, George H. Weiss, and Shlomo Havlin (1989). "Transport Properties of the Continuous-Time Random Walk with a Long-Tailed Waiting-Time Density". In: *Journal of Statistical Physics* 57.1, pp. 301–317.

Werven, Folkert J. van et al. (2009). "Distinct Promoter Dynamics of the Basal Transcription Factor TBP across the Yeast Genome". In: *Nature Structural & Molecular Biology* 16.10, pp. 1043–1048. ISSN: 1545-9985. DOI: 10.1038/nsmb.1674. URL: https://www.nature.com/articles/nsmb.1674 (visited on 09/18/2018).

Widom, J (1985). "Structure of the 3000Å Chromatin Filament: X-Ray Diffraction from Oriented Samples". In: *Cell* 43.1, pp. 207–213. ISSN: 00928674. DOI: 10.1016/0092-8674(85)90025-X. URL: https://linkinghub.elsevier.com/retrieve/pii/009286748590025X (visited on 04/06/2019).

Wiggins, Paul A. et al. (2006). "High Flexibility of DNA on Short Length Scales Probed by Atomic Force Microscopy". In: *Nature Nanotechnology* 1.2. 00290, pp. 137–141. ISSN: 1748-3387. DOI: 10.1038/nnano.2006.63. URL: http://www.nature.com/nnano/journal/v1/n2/full/nnano.2006.63.html (visited on 01/19/2017).

Wilson, A. (2004). "C-Myc Controls the Balance between Hematopoietic Stem Cell Self-Renewal and Differentiation". In: *Genes & Development* 18.22, pp. 2747–2763. ISSN: 0890-9369. DOI: 10.1101/gad.313104. URL: http://www.genesdev.org/cgi/doi/10.1101/gad.313104 (visited on 08/23/2016).

Wong, Hua et al. (2012). "A Predictive Computational Model of the Dynamic 3D Interphase Yeast Nucleus". In: *Current Biology* 22.20, pp. 1881–1890. ISSN: 09609822. DOI: 10.1016/j.cub.2012.07.069. URL: https://linkinghub.elsevier.com/retrieve/pii/S096098221200927X (visited on 04/06/2019).

Woodfin, Beulah M. (1967). "Substrate-Induced Dissociation of Rabbit Muscle Aldolase into Active Subunits". In: *Biochemical and Biophysical Research Communications* 29.3, pp. 288–293. ISSN: 0006291X. DOI: 10.1016/0006-291X(67)90450-0. URL: https://linkinghub.elsevier.com/retrieve/pii/0006291X67904500 (visited on 04/25/2019).

Woringer, Maxime (2014). "Superresolution Insights into the Nuclear Localization of the C-Myc Oncogene". Masters thesis. Paris: École normale supérieure.

Woringer, Maxime and Xavier Darzacq (2018). "Protein Motion in the Nucleus: From Anomalous Diffusion to Weak Interactions". In: *Biochemical Society Transactions*, BST20170310. ISSN: 0300-5127, 1470-8752. DOI: 10.1042/BST20170310. URL: http://www.biochemsoctrans.org/content/early/2018/07/29/BST20170310 (visited on 08/01/2018).

Woringer, Maxime, Xavier Darzacq, and Ignacio Izeddin (2014). "Geometry of the Nucleus: A Perspective on Gene Expression Regulation". In: *Current Opinion in Chemical Biology* 20. 00022, pp. 112–119. ISSN: 13675931. DOI: 10.1016/j.cbpa.2014.05.009. URL: http://linkinghub.elsevier.com/retrieve/pii/S1367593114000647 (visited on 04/04/2017).

Wu, Youjun et al. (2018). "Rapid Diffusion-State Switching Underlies Stable Cytoplasmic Gradients in the Caenorhabditis Elegans Zygote". In: *Proceedings of the National Academy of Sciences* 115.36, E8440–E8449. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1722162115. pmid: 30042214. URL: http://www.pnas.org/content/115/36/E8440 (visited on 09/05/2018).

Wunderlich, Zeba and Leonid A. Mirny (2008). "Spatial Effects on the Speed and Reliability of Protein–DNA Search". In: *Nucleic Acids Research* 36.11, pp. 3570–3578. ISSN: 1362-4962, 0305-1048. DOI: 10.1093/nar/gkn173. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn173 (visited on 10/29/2017).

— (2009). "Different Gene Regulation Strategies Revealed by Analysis of Binding Motifs". In: *Trends in Genetics* 25.10. 00015, pp. 434–440. ISSN: 01689525.

Wutz, Gordana et al. (2017). "Topologically Associating Domains and Chromatin Loops Depend on Cohesin and Are Regulated by CTCF, WAPL, and PDS5 Proteins". In: *The EMBO Journal* 36.24, pp. 3573–3599. ISSN: 0261-4189, 1460-2075. DOI: 10.15252/embj.201798004. pmid: 29217591. URL: http://emboj.embopress.org/content/36/24/3573 (visited on 05/17/2018).

Xiang, Limin et al. (2019). "Super-Resolution Displacement Mapping of Unbound Single Molecules Reveals Nanoscale Heterogeneities in Intracellular Diffusivity". In: *bioRxiv*, p. 559484. DOI: 10.1101/559484. URL: https://www.biorxiv.org/content/10.1101/559484v1 (visited on 03/04/2019).

Xiang, Wanqing et al. (2018). "Correlative Live and Super-Resolution Imaging Reveals the Dynamic Structure of Replication Domains". In: *bioRxiv*, p. 189373. DOI: 10.1101/189373. URL: https://www.biorxiv.org/content/early/2018/01/24/189373 (visited on 02/07/2018).

Yan, Jian et al. (2013). "Transcription Factor Binding in Human Cells Occurs in Dense Clusters Formed around Cohesin Anchor Sites". In: *Cell* 154.4. 00125, pp. 801–813. ISSN: 00928674. DOI: 10.1016/j.cell.2013.07.034. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867413009422 (visited on 02/07/2017).

Yan, Jie et al. (2007). "Micromanipulation Studies of Chromatin Fibers in *Xenopus* Egg Extracts Reveal ATP-Dependent Chromatin Assembly Dynamics". In: *Molecular Biology of the Cell* 18.2. Ed. by Kerry Bloom, pp. 464–474. ISSN: 1059-1524, 1939-4586. DOI: 10.1091/mbc.e06-09-0800. URL: http://www.molbiolcell.org/doi/10.1091/mbc.e06-09-0800 (visited on 04/06/2019).

Yang, Si Kyung et al. (2011). "Monovalent, Clickable, Uncharged, Water-Soluble Perylenediimide-Cored Dendrimers for Target-Specific Fluorescent Biolabeling". In: *Journal of the American Chemical Society* 133.26, pp. 9964–9967. ISSN: 0002-7863. DOI: 10.1021/ja2009136. URL: http://dx.doi.org/10.1021/ja2009136 (visited on 10/29/2017).

Yeh, Elizabeth et al. (2004). "A Signalling Pathway Controlling C-Myc Degradation That Impacts Oncogenic Transformation of Human Cells". In: *Nature Cell Biology* 6.4, pp. 308–318. ISSN: 1465-7392, 1476-4679. DOI: 10.1038/ncb1110. URL: http://www.nature.com/doifinder/10.1038/ncb1110 (visited on 11/18/2017).

Yeo, Jia-Chi and Huck-Hui Ng (2013). "The Transcriptional Regulation of Pluripotency". In: *Cell research* 23.1. 00046, pp. 20–32. URL: http://www.nature.com/cr/journal/v23/n1/abs/cr2012172a.html (visited on 09/29/2016).

Yi, Xiyu, Rafael Piestun, and Shimon Weiss (2019). "3D Super-Resolution Imaging Using a Generalized and Scalable Progressive Refinement Method on Sparse Recovery (PRIS)". In: *bioRxiv*. DOI: 10.1101/532143. URL: http://biorxiv.org/lookup/doi/10.1101/532143 (visited on 02/22/2019).

Yochum, Gregory S. (2011). "Multiple Wnt/ß-Catenin Responsive Enhancers Align with the MYC Promoter through Long-Range Chromatin Loops". In: *PLoS ONE* 6.4. Ed. by Mary Bryk, e18966. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0018966. URL: http://dx.plos.org/10.1371/journal.pone.0018966 (visited on 09/07/2017).

Youmans, Daniel T., Jens C. Schmidt, and Thomas R. Cech (2018). "Live-Cell Imaging Reveals the Dynamics of PRC2 and Recruitment to Chromatin by SUZ12-Associated Subunits". In: *Genes & Development*. ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.311936.118. pmid: 29891558. URL: http://genesdev.cshlp.org/content/early/2018/06/11/gad.311936.118 (visited on 07/09/2018).

Yu, Chen et al. (2016). "Structure-Based Inhibitor Design for the Intrinsically Disordered Protein c-Myc". In: *Scientific Reports* 6, srep22298. ISSN: 2045-2322. DOI: 10.1038/srep22298. URL: https://www.nature.com/articles/srep22298 (visited on 11/09/2017).

Zada, AA et al. (2006). "Proteomic Discovery of Max as a Novel Interacting Partner of C/EBPa: A Myc/Max/Mad Link". In: *Leukemia* 20. 00000, pp. 2137–2146.

Zaidi, Hussain A., David T. Auble, and Stefan Bekiranov (2017). "RNA Synthesis Is Associated with Multiple TBP-Chromatin Binding Events". In: *Scientific Reports* 7, p. 39631. ISSN: 2045-2322. DOI: 10.1038/srep39631. URL: http://www.nature.com/articles/srep39631 (visited on 11/18/2017).

Zaret, Kenneth S. et al. (2008). "Pioneer Factors, Genetic Competence, and Inductive Signaling: Programming Liver and Pancreas Progenitors from the Endoderm". In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 73. 00112. Cold Spring Harbor Laboratory Press, pp. 119–126. URL: http://symposium.cshlp.org/content/73/119.short (visited on 12/23/2016).

Zeller, Karen I., Anil G. Jegga, et al. (2003). "An Integrated Database of Genes Responsive to the Myc Oncogenic Transcription Factor: Identification of Direct Genomic Targets". In: *Genome Biology* 4.10, R69. ISSN: 1474-760X. DOI: 10.1186/gb-2003-4-10-r69. URL: https://doi.org/10.1186/gb-2003-4-10-r69 (visited on 03/31/2019).

Zeller, Karen I., XiaoDong Zhao, et al. (2006). "Global Mapping of C-Myc Binding Sites and Target Gene Networks in Human B Cells". In: *Proceedings of the National Academy of Sciences* 103.47, pp. 17834–17839. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0604129103. pmid: 17093053. URL: http://www.pnas.org/content/103/47/17834 (visited on 11/21/2018).

Zhan, Yinxiu et al. (2017). "Reciprocal Insulation Analysis of Hi-C Data Shows That TADs Represent a Functionally but Not Structurally Privileged Scale in the Hierarchical Folding of Chromosomes". In: *Genome Research* 27.3, pp. 479–490. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.212803.116. URL: http://genome.cshlp.org/lookup/doi/10.1101/gr.212803.116 (visited on 04/06/2019).

Zhang, Yifei et al. (2018). "Aldolase Does Not Show Enhanced Diffusion in Dynamic Light Scattering Experiments". In: *Nano Letters* 18.12, pp. 8025–8029. ISSN: 1530-6984, 1530-6992. DOI: 10.1021/acs.nanolett.8b04240. URL: http://pubs.acs.org/doi/10.1021/acs.nanolett.8b04240 (visited on 04/25/2019).

Zhang, Z. et al. (2011). "A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome". In: *Science* 332.6032, pp. 977–980. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1200508. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1200508 (visited on 04/06/2019).

Zhao, Xi, Krishna K. Dey, et al. (2017). "Enhanced Diffusion of Passive Tracers in Active Enzyme Solutions". In: *Nano Letters* 17.8, pp. 4807–4812. ISSN: 1530-6984, 1530-6992. DOI: 10.1021/acs.nanolett.7b01618. URL: http://pubs.acs.org/doi/10.1021/acs.nanolett.7b01618 (visited on 04/25/2019).

Zhao, Xi, Henri Palacci, et al. (2018). "Substrate-Driven Chemotactic Assembly in an Enzyme Cascade". In: *Nature Chemistry* 10.3, pp. 311–317. ISSN: 1755-4330, 1755-4349. DOI: 10.1038/nchem.2905. URL: http://www.nature.com/articles/nchem.2905 (visited on 04/25/2019).

Zhao, Z. W. et al. (2014). "Spatial Organization of RNA Polymerase II inside a Mammalian Cell Nucleus Revealed by Reflected Light-Sheet Superresolution Microscopy". In: *Proceedings of the National Academy of Sciences* 111.2. 00048, pp. 681–686. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1318496111. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1318496111 (visited on 11/18/2016).

Zhen, Chao Yu et al. (2016). "Live-Cell Single-Molecule Tracking Reveals Co-Recognition of H3K27me3 and DNA Targets Polycomb Cbx7-PRC1 to Chromatin". In: *eLife* 5, e17667. ISSN: 2050-084X. DOI: 10.7554/eLife.17667. URL: https://elifesciences.org/articles/17667 (visited on 09/25/2017).

Zhou, Feng et al. (2013). "Genome-Scale Proteome Quantification by DEEP SEQ Mass Spectrometry". In: *Nature Communications* 4. 00039. ISSN: 2041-1723. DOI: 10.1038/ncomms3171. URL: http://www.nature.com/doifinder/10.1038/ncomms3171 (visited on 04/26/2017).

Zhu, Lian and Clifford P Brangwynne (2015). "Nuclear Bodies: The Emerging Biophysics of Nucleoplasmic Phases". In: *Current Opinion in Cell Biology* 34, pp. 23–30. ISSN: 09550674. DOI: 10.1016/j.ceb.2015.04.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0955067415000435 (visited on 04/06/2019).

Zuin, Jessica et al. (2014). "A Cohesin-Independent Role for NIPBL at Promoters Provides Insights in CdLS". In: *PLoS Genetics* 10.2. Ed. by Gregory S. Barsh, e1004153. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1004153. URL: http://dx.plos.org/10.1371/journal.pgen.1004153 (visited on 04/06/2019).

Zwier, Matthew C. et al. (2016). "Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein–Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered P53 Peptide". In: *The Journal of Physical Chemistry Letters* 7.17, pp. 3440–3445. ISSN: 1948-7185. DOI: 10.1021/acs.jpclett.6b01502. URL: http://pubs.acs.org/doi/10.1021/acs.jpclett.6b01502 (visited on 02/05/2018).