bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Polymer physics of nuclear organization and function

A. Amitai[1] and D. Holcman [2,3] *

Abstract

We review here recent progress to link the nuclear organization to its function, based on elementary physical processes such as diffusion, polymer dynamics of DNA, chromatin and the search mechanism for a small target by double-stranded DNA (dsDNA) break. These physical models and their analysis make it possible to compute critical rates involved in cell reorganization timing, which depends on many parameters. In the framework of polymer models, various empirical observations are interpreted as anomalous diffusion of chromatin at various time scales. The reviewed theoretical approaches offer a framework for extracting features, biophysical parameters, predictions, and so on, based on a large variety of experimental data, such as chromosomal capture data, single particle trajectories, and more. Combining theoretical approaches with live cell microscopy data should unveil some of the still unexplained behavior of the nucleus in carrying out some of its key function involved in survival, DNA repair or gene activation.

Keywords: polymer models, nucleus organization, chromatin, modeling, stochastic processes, first passage times, search time, narrow escape theory, single particle trajectories, anomalous diffusion, looping, eigenvalue expansion, Fokker-Planck equation, asymptotic formula, boundary layer, chromosomal capture data, Hi-C analysis, diffusion processes, tethering force, potential wells, stochastic simulations, coarse-grained analysis, mean square displacement, telomeres, clustering, aggregation, DNA repair, homologous recombination.

Contents

1 Introduction **5**

2 A brief description of the nuclear organization **11**

2.1 Eukaryotic chromatin 11
2.2 Organization of telomeres 11
2.3 DNA organization in bacteria 12
2.4 Chromosomal territories 13
2.5 Spontaneous DNA loops 14
2.6 Motion of a single DNA locus 14
2.7 Chromosome conformation capture data 14
2.8 Spatial nuclear organization for gene regulation, DNA maintenance and repair 16

*[1] IMES, MIT, Cambridge, MA, USA. [2] Department of Applied Mathematics and Theoretical physics (DAMTP), University of Cambridge, CB30DS, Cambridge, United Kingdom and [3] Ecole Normale Superieure, Paris, France.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3 Polymer models for modeling DNA dynamics **17**
3.1 Elementary polymer chains . . . 18
3.2 Polymer with self-avoiding interactions in confined domains . . . 19
3.3 The Smoluchowski limit of the Langevin equation for polymer dynamics . . . 20
3.4 The Rouse model . . . 21
3.5 β-polymer as a generalized Rouse model . . . 22
3.6 Numerical implementation of polymer simulation using Euler's scheme . . . 24
3.7 Implementing excluded volume interactions . . . 24
3.8 Modeling stiffness . . . 25
3.9 Rod-like model . . . 27
3.10 Anomalous diffusion in polymer models . . . 27
3.10.1 Normal and anomalous diffusion regimes for a single monomer . . . 30
3.10.2 Anomalous motion of a Rouse polymer . . . 30
3.10.3 Monomer motion characterized by a power-law decay time . . . 31
3.10.4 Correlation function for a Rouse and $\beta-$polymer . . . 32

4 Polymer looping and the search for a small target: a mean passage time study **33**
4.1 Brief analytical formulation . . . 35
4.2 Szabo, Schulten& Schulten approach and later analysis . . . 35
4.3 Closed loop configuration space for a rod-like polymer . . . 37
4.4 Looping for a Rouse polymer . . . 37
4.5 Computing the eigenvalues of the Fokker-Planck equation and the MFET . . . 39
4.6 Looping in a confined domain . . . 43
4.7 Detail computations of the looping time in a confined domain . . . 45
4.8 Brownian simulations of looping in a confined domain . . . 48
4.9 Looping formula for rod-like long polymer . . . 48
4.10 Looping time for a $4^{th}-$order stiff polymer using a non-Markovian approach . . . 50
4.11 Mean First Encounter Time (MFET) between two monomers located on the same $\beta-$polymer . . . 51
4.12 Search for a small target located on a boundary . . . 53
4.13 Induced screening potential on a monomer by the other ones at a boundary . . . 56
4.14 NETP with additional stiffness . . . 59
4.15 Summary of the looping time formulas . . . 59
4.15.1 Mean First Encounter Time formula in free space for a Rouse polymer . . . 59
4.15.2 Mean First Encounter Time formula in a confined domain for a Rouse polymer 61
4.15.3 Mean First Encounter Time formula for a β-polymer. . . . 62
4.15.4 Arrival time of a monomer to a small target located on the boundary of a bounded domain . . . 62

5 Analyzing chromatin organization and dynamics using polymer model **63**
5.1 Numerical simulations and chromatin dynamics . . . 63
5.2 Characterizing chromosomes and nucleus architecture using large scales simulations . . 65
5.3 Interpreting chromosome capture data inside nuclear territories from polymer looping 67
5.4 Interpreting CC-data using polymer models . . . 69
5.4.1 Fractal polymer globule . . . 69
5.4.2 Interpreting CC-data using the $\beta-$polymer model . . . 69
5.4.3 String binder and switch model (SBS) model . . . 69

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5.5 Using CC-data and encounter probabilities to describe nuclear sub-confinement 70

6 Chromatin modeling using polymer 73
6.1 Chromatin modeled by co-polymer 73
6.2 Chromatin condensation/aggregation depending on chromatin state 74
6.3 Constructing an effective polymer model from Hi-C maps to sample the space of chromatin configuration 74
6.4 Chromatin loops extrusion as a mechanism for chromosome segregation and TADs . . 77

7 Coarse-graining polymer models for analyzing telomere organization in yeast nucleus 82
7.1 Encounter rate of telomeres 83
7.2 Dissociation of a telomere from a cluster 85
7.3 Characterization of telomere exchanged between clusters 85

8 Single particle trajectories of a chromatin locus 87
8.1 Live imaging of nuclear elements 87
8.2 Statistics analysis of SPTs 87
8.3 Second moment statistics of SPTs 90
8.3.1 Constraint length of a locus 91
8.3.2 Effective diffusion coefficient of a chromatin locus 91
8.4 Anomalous diffusion of a chromatin locus 91
8.4.1 Variability of the anomalous exponent over yeast cell population 91
8.4.2 Increasing the anomalous exponent by applying an external drift 93
8.4.3 Oscillating forces acting on a monomer modify the anomalous exponent 94
8.5 Probing chromosomal structure and dynamics in bacteria using SPTs 94
8.6 Local interactions acting on chromatin 95
8.6.1 Recovering a tethering force (external potential) for a Rouse polymer 98
8.6.2 Extracting a deterministic force from an ensemble observed locus trajectories . 99
8.6.3 Polymer dynamics constricted by two tethering forces 100
8.6.4 Analyzing tethering forces for a β-polymer 101
8.6.5 Extracting forces acting on a locus from SPTs 101
8.6.6 Recovering applied forces from the auto-correlation function 103
8.6.7 Distribution of forces of MAT-locus from live cell imaging SPT in yeast 103

9 Physical principles underlying DNA Repair 104
9.1 Encounter loci at the nuclear envelope 105
9.2 Physical constraints underlying double-stranded DNA break repair 105
9.3 Increasing DSB dynamics reduces the MFET 107
9.4 DSB re-localization as a protection mechanism 109

10 Conclusion and perspective 111
10.1 Which polymer model to use and when? 111
10.2 Possible directions and open questions 112
10.2.1 Extracting forces from multiples tagged loci 112
10.2.2 Is there a best model to describe Hi-C data and the encounter probability matrix? 112
10.2.3 How to reconciliate SPTs and Hi-C statistics? 112
10.2.4 Proteins remodeling during DNA break repair 113

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10.2.5 Revisiting the classical question of the search process for a promoter site by a transcription factor . . . 113
10.2.6 Computing the probability of a translocation during Non-Homologous End-Joining (NHEJ) . . . 113
10.2.7 Modeling the random initiation of gene factories . . . 113
10.2.8 Nucleosome repositioning following DNA damages . . . 114
10.2.9 Limitation of statistical analysis for High C-data . . . 114
10.2.10 Can statistical analysis of Hi-C data reveal novel features hidden in the genome organization of cancer cells or cells infected by viruses . . . 114

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Introduction

The eukaryotic cell nucleus separates the genetic material carried by the DNA molecules from the rest of the organelles as opposed to prokaryotes where there is no membrane separation [5]. The cell nucleus is certainly the ultimate key subcellular compartment to understand, because its organization is essential for a myriad of DNA based transactions such as transcription or DNA repair. Understanding its complexity is a joint effort of both theoretical and experimental approaches. We review here recent polymer models and their analysis, computational methods and numerical simulations, statistical and data analysis approaches for extracting information from large amounts of experimental data and predicting the main driving forces responsible for nuclear organization and function.

Physical models driven by diffusion at different length/time scales [119] and polymer models can be used to describe DNA and the chromatin molecules [226, 153], aggregation-dissociation with a finite number of particles [123] or the time for Brownian particle to find small targets [55]. The mathematical analysis of these models is used to estimate short- and long-time behavior of rare events, such as molecular binding, DNA looping, DNA repositing to the small location or DNA search for homology. The asymptotic analysis of the model equation reveals how key physical parameters control the spatial scales of the nucleus starting at a molecular level.

We recall that the nucleus contains the genetic material folded in multiscale but yet unknown or-

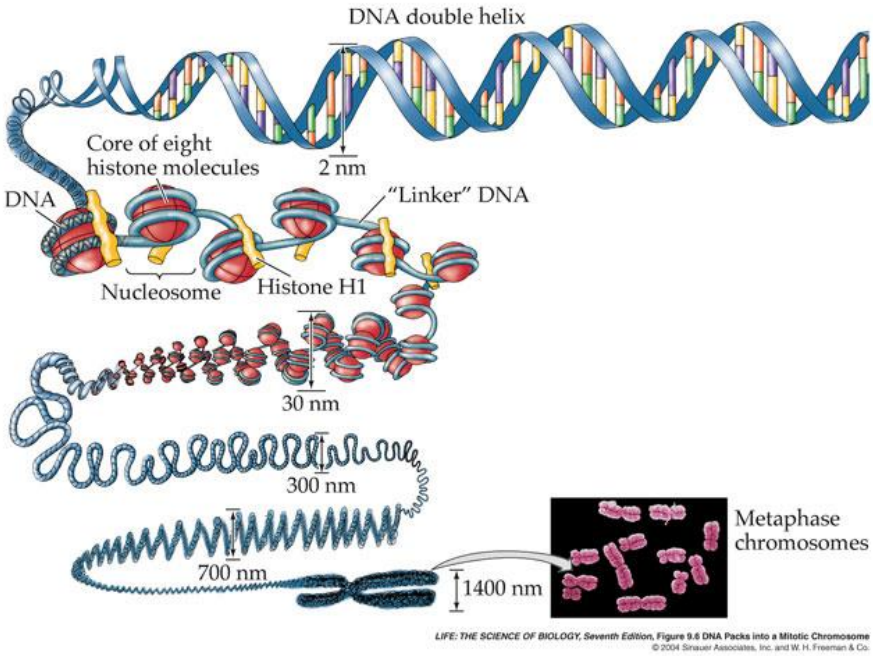


Fig. 1: **Schematic representation of the multiscale chromatin fiber and chromosome structure (adopted from [210]).**

ganization (Fig. 1). Genetics assays that introduce mutations to the DNA can perturb its function and change the nucleus organization. While these local changes are often difficult to resolve spatially, they can however be studied using polymer models and analyzed by stochastic processes [89].

The multiscale organization starts at the molecular level (the four bases A,C,T,G). Eukaryotic DNA is wrapped around histone proteins, mediating mechanical and electrostatic interactions, to

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

form a unit called a nucleosome (Fig. 1) [165]. The spatial scale above the nucleosome diameter is often referred to as *beads-on-a- string*, where nucleosomes form beads connected by linkers. Nucleosomes are further compacted by different proteins such as linker histones to form another structure called the *30 nanometer fiber* [77]. The fiber is organized in a hierarchical structure whose nature is now being revealed, until the very coarse-grained level of a whole chromosome [80] (Fig. 1). Integrating all these interactions from molecular to the chromatin level is certainly a challenge for modeling.

While the 10 nm level scale is supported by the nucleosome on which the DNA is wrapped, the organization and folding at higher scales is still under debate. This multiscale representation has also been recently challenged because the nucleus is highly dynamic and chromosomes are constantly moving, opening or reorganizing [59, 199], an activity that can be directly monitored using live cell microscopy at an unprecedented spatio-temporal resolution of tens of nanometers and milliseconds [243, 155].

Live-cell imaging of the chromatin and polymer models are now converging. If chromatin locus can now be routinely tracked at a spatial scale of tens of nanometer resolution and at a time resolution of hundreds of milliseconds or less [40, 172, 25], in parallel, polymer models and their numerical simulations are constantly increasing their accuracy by incorporating physical and mechanical properties of the DNA or the chromatin structure to reproduce their local organization [173, 222, 187, 223, 272, 291, 274, 10].

Statistical quantities derived from Brownian motion theory [39, 69, 181], such as the Mean-Square-Displacement (MSD), computed along single particle trajectories are used to analyzed chromatin properties. The MSD is used as an indicator of spatial confinement [170, 111, 68], but it remains difficult to give an final interpretation, because the motion of a single DNA locus cannot be reduced to Brownian motion. The motion is indeed much more complex, because a locus is a bead moving with the entire polymer. Thus polymer models are needed to interpret the statistical analysis and bridge the gap between the output of statistical analysis obtained for a single locus and the reconstruction of the underlying chromatin organization. Additional polymer properties are clearly necessary to interpret the motion of a single DNA locus, especially when the entire DNA molecule is moving in a confined environment, or when a small promoter or DNA site is searching for a specific partner that can be relocated on the nuclear surface (Fig. 3). In most cases, statistical analysis has so far revealed deviation from diffusion called anomalous diffusion, for which the interpretation still remains unclear.

The effort of revealing the nucleus organization does not simply prevail for mammalian cells, the chromosomal architecture in bacteria is quite complicated and appear also to be modulated both spatially and temporally. High-resolution single particle trajectories have revealed rare but rapid ballistic trajectories. While the origin of this motion is still unclear, polymer models are used to examine whether subdiffusive chromosomal dynamics exhibits these rapid movements or they are due to external forces (see [129, 128] and [175] for a general review on physical interpretation of bacterial DNA based on sub-diffusion theory.)

Another difficulty inherent to the cell nucleus is the absence of clear main driving forces. In physiological conditions, ionic concentrations vary from tens to hundreds millimolar, and thus an average electrostatic force cannot act at a distance larger than few nanometers, as measured by the Debye length [58]. This electrostatic force has a fast decay due to a screening potential induced by the statistical properties of ions in solution. Consequently, two charged molecules cannot directly interact unless they are moved in close proximity, a process mostly driven by diffusion. Thus, electrostatic force should not have any role in long-range forces in the nucleus. Another source of chromatin reorganization and motion is due to actin fibers in the cytoplasm outside the nucleus, that is physically connected to the nucleus membrane and can affect directly the nucleus motion.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

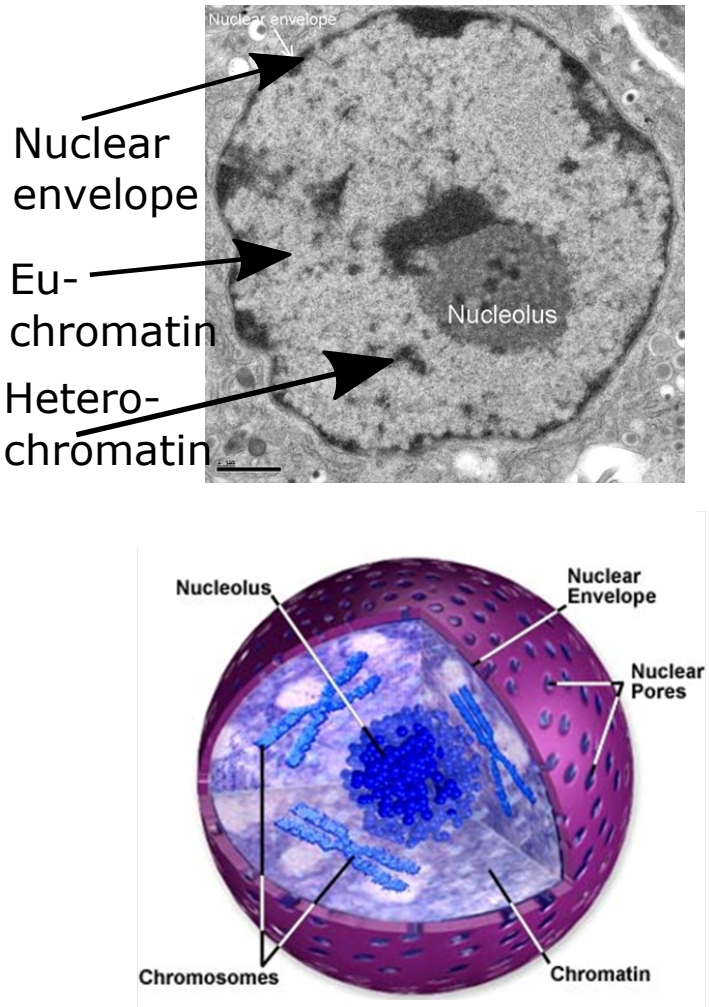


Fig. 2: **Organization in the cell nucleus. Upper:** Electron-Microscopy image of the nucleus. We have indicated the different region: heterochromatin(dense black regions) and euchromatin (white region) and the nuclear boundary **Lower:** artistic representation of decondensed chromosomes (blue) floating in the nucleus. Nuclear pores are located on the surface.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The genetic material organization seems to be dominated by random forces at various scales, but it is unclear how fast and how far these local forces can affect the long-range chromatin organization. In addition, it is also unclear how the local chromatin loci are affected by diffusion. Local diffusion is generated by random collisions, but the overall random motion also account for transient interactions with surrounding molecules and thus does not simply refer to classical Brownian motion of a particle. Random motion of a locus probably includes random forces that originate from the thermal noise in a crowded and organized structures such as the chromatin fiber. This subject is not new in stochastic processes, where models are used to study correlated noise and how random collisions influence large molecular structures [236, 238, 237]. Complex noise can also be characterized statistically using the power density spectrum [237] for single particle trajectories, but here again the interpretation of the data should be based on the statistical properties of polymer models. In summary, chromatin is constantly moving driven by various cellular sources leading to random fluctuations inside confined nuclear sub-domains and restricted by the nuclear envelope in eukaryotes. Accounting for these forces is a clear challenge for polymer physics and for the interpretation of data.

In this review, we present several polymer models to account for the physical forces underlying nuclear organization. We shall first summarize the classical polymer models (section 3) followed by an analysis of polymer encounter events that are responsible for generating the large spectrum of time scales involved in nucleus organization (section 4). Indeed, the intrinsic polymer structure is certainly involved in the genesis of long-time correlation between the different spatial scales. Thus, the description of a monomer or a polymer dynamics does not fall into the classical random walk or the idealized Brownian mathematical framework. In addition are events arising from stochastic polymer processes such as random looping, search for a small target or long-range reorganization is probably the appropriate framework to analyze chromatin organization. A classical example is polymer looping [202, 209] (Fig. 14), where two sites are brought together for possible interaction. Analysis and simulations permit computing the rate of looping, which is the reciprocal of the mean first encounter time (MFET) between two monomers. Such rate provides great constraints on polymer dynamics and predicts the frequencies of encounter between two given loci on the chromatin. The analytical expression of these rates constitutes the emerging polymer physical laws for DNA and chromatin, that clarifies the dependency of many small and large parameters, especially when it becomes necessary to interpret the motion of a single tagged locus located on a (non labeled or hidden) polymer and to extract biophysical parameters (section 6). Polymer models can be used to analyze anomalous diffusion of a single locus, characterized by a parameter α, which is the exponent of the time-correlated function for small time, approximated by $\approx t^{\alpha}$ [104, 34, 275, 284]. As we shall see, this anomalous behavior is generated by the collective motion of the DNA, again captured by polymer models and reviewed in details in section 3.

The chromatin is constantly remodeled, during cell cycle or under different stress conditions [255], preventing the nucleus from reaching a steady-state equilibrium. Telomeres which constitute the chromosome ends (fig. 5) can associate and dissociate in small cluster dynamically. Chromosome sites do encounter with different frequency depending on their locations, while gene regulation and DNA repair are possible based on constant re-organization of the chromatin [67, 265, 159]. These encounter events can be characterized by their frequency distribution that we review in section 5. Interacting sites coming into proximity is a random and rare event, with a long time scale compared to single particle diffusion [119]. The exact implications of these encounter events are still unknown, but over the past decades, many experimental methods have been developed. One critical and very prolific was the chromosome capture method [61], from which the encounter probability of DNA sites is extracted and the contact frequency can be estimated [83], as discussed in sections 5 and 6.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A down point of the theoretical approach is the actual performance of numerical algorithms to simulate chromatin dynamics at an atomic level. These performances are still insufficient to reconcile them with the need of following the nucleus reorganization over minutes to hours, as it would involve simulating too many components. Simulating the encounter between DNA-sites located far away is still hard because they are rare events, requiring long simulations and asymptotic formula (see section 4). But analytical and coarse-graining modeling methods are valuable tools especially to estimate fundamental time scales involved in encounter processes of macromolecules or during DNA repair (section 9). They are also used to extract biophysical properties of single particle trajectories of chromatin locus using live cell microscopy approaches [265, 155] (section 8). Interpreting data produced by these dynamical methods continue to benefit from polymer models, but reconstructing the chromosomal organization from these data remains a challenging hurdle.

In summary, this review aims to describe polymer physical models, numerical simulations and data

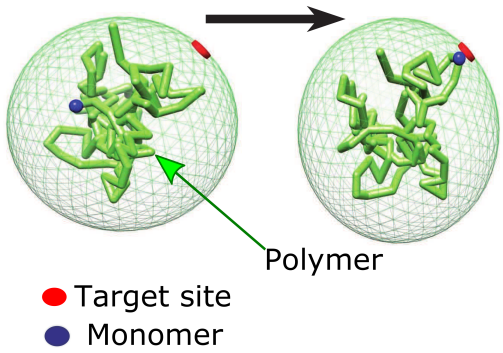


Fig. 3: **Schematic representation of a polymer.** A monomer (blue) encounters after some time a small target (red) located on the nuclear boundary (figure from the group).

analysis in order to understand chromatin organization in the nucleus of mammalian, yeast cells or even inside bacteria. It is organized as followed. Part 1 (section 2) is a very short description of the cell nucleus organization, summarizing basic properties of physical modeling. Current methods employed to explore the nucleus include live cell imaging, single particle trajectories approaches and more recently chromosome conformation capture methods, producing million by million matrices of encounters between DNA pairs. The second part (section 3 and 4) describes polymer models, starting with Rouse and other associated polymer models used to described chromatin. Asymptotic methods are used to derive formulas for the encounter time of two monomers during DNA-looping in a free and a confinement domain. These formulas reveal the structure of the parameter space and the role of singular parameters, which are usually the bottle necks of long numerical simulations and probably the underlying biophysical processes they represent. These analytical methods are based

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

on eigenvalues expansion of the Fokker-Planck equation in a high dimensional space (the dimension of which is the number of beads of the polymer). The associated numerical simulations are based on stochastic dynamics, also used to reconstruct the chromosomal organization in the nucleus and to interpret chromosome capture data (section 5). In section 6, we discuss some modeling approach of the chromatin as a heterogeneous strand, aimed at interpreting the formation of topological domains in chromosomes and the relation between the epigenetic state of the chromatin and its structure. We describe in chapter 7 the encounter dynamics and organization of yeast telomere. In section 8, we summarize various approaches where polymer models are combined with live cell imaging data to unveil physical properties, such as normal and anomalous diffusion of a single chromatin locus trajectories, depending on the time scale of experiments. We mention several criteria to identify the appropriate time scale and nature of the underlying motion. Section 9 discusses applications of polymer models to various scenarios underlying major events in the nucleus, such as double-stranded DNA break repair, which involves a still unclear search process for a homologous sequence. In the final part 10, we summarize the role of polymer physics in interpreting the chromatin organization and discuss open questions.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2 A brief description of the nuclear organization

The nucleus of eukaryotic cells contains chromosomes and many other components such as nuclear pores and the nucleolus, a subcompartment composed of proteins and Ribosomal RNAs (rRNAs) [52] (Fig. 2, references are *www.lifesci.sussex.ac.uk/home/Julian_Thorpe/tem29.jpg* see also *vle.du.ac.in/mod/book/print.php?id = 9345#ch13971* and *www.pinterest.com/pin/83316661827063673/*) and Fig. 4). These components are implicated in proteins production, mRNA regulation and cell homeostasis. Yet, how these elements are organized in the nucleus remains poorly understood. The nucleus is not segregated in expressed and non-expressed genes, but the organization is certainly involved in the regulation of gene expression and repair mechanisms.

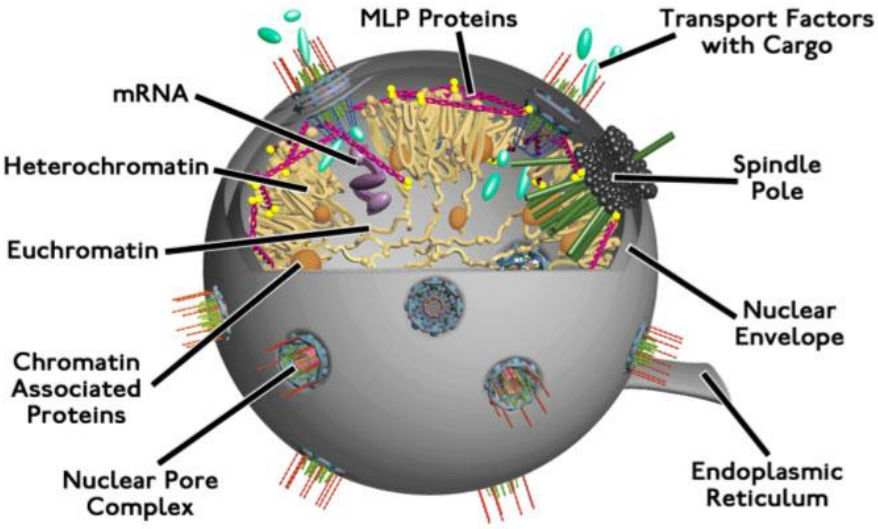


Fig. 4: **Schematic representation of the yeast nucleus with various components (Rockefeller University (http://lab.rockefeller.edu/rout/resproj3).**

2.1 Eukaryotic chromatin

The DNA is usually covered with an ensemble of molecules, forming the chromatin, which has an intrinsic structure, imposed by cylindrical protein complexes called nucleosomes [209, 165, 77] (see also [53] for a recent review on the physics of epigenetics). Beyond a 30nm scale, the chromosome structure is thought to be non-uniformly distributed in the nucleus. In yeast, the chromosomes fold back, called the Rabl configuration [131], resulting from the attachment of microtubules to a structure (the kinetochores) assembled at centromeric chromatin [213]. Microtubules are connected to a structure called the spindle pole body [265] (Fig. 4). In Metazoan, the nucleus is more dense (Fig. 6) and chromosomes adopt a globule shape. Interestingly, during cell division chromosomes are pulled by microtubules through the kinetochores into daughter cells.

2.2 Organization of telomeres

The ends of the chromosomes are made of repetitive sequences, the size of which changes after each cell division [27, 87]. In yeast, telomeres are located on the nuclear periphery and form stable

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

but transient clusters through the interactions of proteins (Sir4 and yKu70/yKu80 proteins) [264, 265, 225, 124] with other proteins of the nuclear envelope such as Mps3 (Fig. 5). Telomeres are not uniformly distributed at the nuclear periphery, but form small dynamical foci [94] and their number and composition evolve over time [225, 124]. These clusters can promote gene silencing [225, 124], depending on (the Sir) proteins and on the spatial position of telomeres. Telomeres tend to cluster according to their chromosomal length, such that telomeres belonging to longer (or shorter) chromosomes cluster together [269]. Telomeres of mammalian cells are distributed more or less evenly over the nucleus. They can form clusters at the periphery [186], but also internally on nuclear bodies.

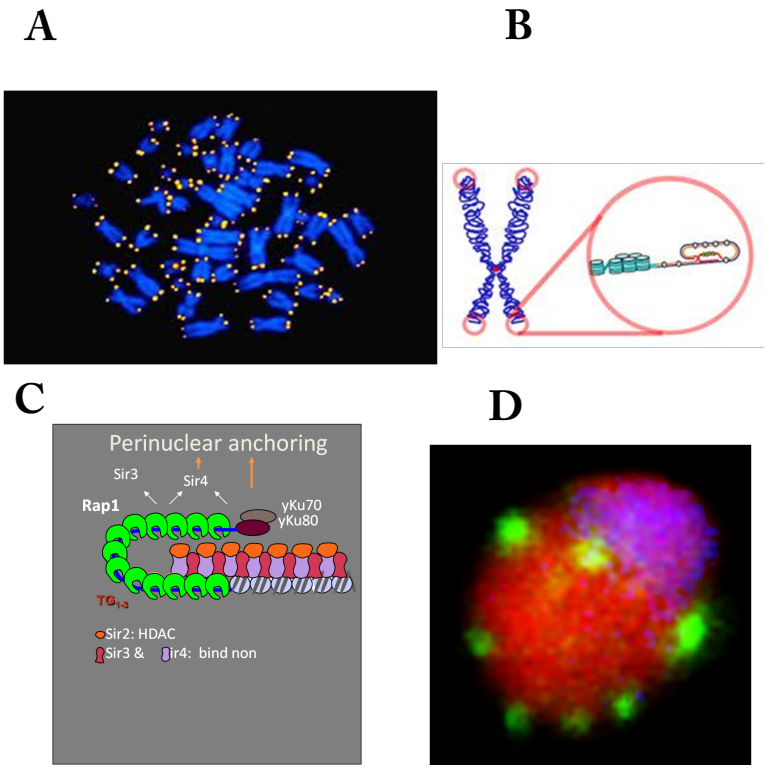


Fig. 5: **Telomeres (located at the end of the chromosomes) in the cell nucleus.** (**A**) Telomere are marks in yellow. (**B**) Magnification of the telomere geometry.(**C**) local molecular organization of telomeres (**D**) Telomeres organized in clusters at the surface of the yeast nucleus [265].

2.3 DNA organization in bacteria

In bacteria, DNA, RNA and proteins are present in the same single cell compartment and the chromosome is a single DNA circular molecule further organized with complement proteins (the nucleoids). Despite the absence of intracellular membranes, bacteria have a high degree of intracellular spatial organization (see [175] for a recent physical and biological review). In the two bacteria E. coli and Caulobacter, fluorescent tagging of chromosomal loci reveals linear spatial arrangement of loci according to their chromosomal coordinate, the presence of chromosomal fiber and a highly non-uniform chromosomal structure, suggesting a compartmentalized structure with boundaries. Four macrodomains of a few hundred kbs in size have been identified, corresponding to regions surrounding

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the replication origin. 200-400 chromosomal domains have been identified in E. coli, characterized by macromolecular crowding, local electrostatic forces, supercoiling and nucleoid proteins are contributing to DNA compaction and organization [175].

2.4 Chromosomal territories

Live cell microscopy data suggest that chromosomes are organized in distinct geographical territories with heterogeneous shapes and sizes [54] (Fig. 6). This organization prevents chromosomes to mix. This organization is not prominent in yeast, probably because the nucleus is not dense enough with chromatin, however it becomes prevalent in higher eukaryotes, where the nucleus is larger and contains a higher density of DNA, generating a crowded environment.

Interestingly, changing chromosomes organization by clustering them together or part of them or several parts of different chromosomes can activate or repress genes. This reorganization could allow the simultaneous regulation of multiple loci in specific substructures, called 'transcription factories', although the existence of these factories is still debated [259, 51]. Gene expression can be modulated by regulatory elements, such as enhancers [209], which are often located far away from their target genes along the DNA strand. Many enhancer proteins are thought to associate with the promoter sites on the DNA molecule that they regulate, but a chromatin loop can also bring a molecule close to its promoter site, as observed using fluorescent markers [166].

Gene position affects transcriptional activity, but there is no general rule across eukaryotes: in yeast,

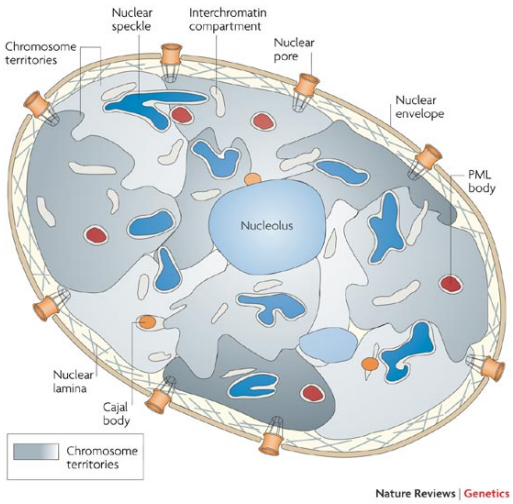


Fig. 6: **Representation of the cell mammalian nucleus organization.** The nuclear envelope contains small pores and intermediate filaments. Nuclear sub-organized regions cluster to form nucleoli, near chromosome territory (gray) and interchromatin compartments. The chromatin is organized in distinct territories (reproduced from [150]).

genes relocate to the nuclear periphery after activation [266]. In mammalian cells, they relocate from the periphery towards the interior of the nucleus once they have been switched on [176]. Similarly, a number of highly transcribed gene-dense regions for example the MHC class II locus are expelled from chromosome territories, once they are activated after the formation of a large chromatin loop [281].

In summary, the organization of the genome is characterized by interactions of chromatin with nuclear sites, influencing gene regulation depending on the location. However, many genes do not

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

change their expression level based on their position and active genes can be found both at the periphery or deep inside a chromosome territory.

2.5 Spontaneous DNA loops

When a DNA molecule bends enough to bring into close proximity two sites located faraway, a loop is formed [209]. Surprisingly, genes can be regulated by random looping events, when a promoter site interacts with an operator sequence. The classical example is the Lac operon system [148], where the lac repressor functions through clamping of two out of three target DNA sites (operators). Between the two sites, the formed loop can interfere with RNA polymerase while reading a gene sequence [174]. This process has been investigated experimentally [66, 76], numerically [47] and analytically [288, 261, 202, 13, 9] revealing the rate at which looping occurs. This rate is the limiting step for many other biological activities. A large part of this review will be dedicated to these looping events.

2.6 Motion of a single DNA locus

The motion of a single DNA locus is monitored using live cell imaging microscopy in bacteria or in eukaryotes and is described as random [265] or directed [111]. However, a unified physical description is still lacking. Interestingly, glucose depletion or the addition of protonophore to deplete the membrane potential, abolishes large range movements [170]. These effects suggest that interphase chromatin movement is sensitive to various ATP energy levels. Chromatin remodeler molecules rather than RNA or DNA polymerases [265] underlie the ATP-driven motions. Moreover, chromatin movement can be modulated depending on the cell phase. For example, it decreases as cells enter in the S-phase, which may be related to the binding of molecules such as cohesin to the chromatin [111, 70].

2.7 Chromosome conformation capture data

Chromatin organization in the nucleus was analyzed by estimating the frequency by which two pairs (small DNA fragments) come into close proximity (fig. 7) [61]. In these experiments, the spatial chromosome position and interactions are detected by literally freezing the nucleus by chemical ligation of near-by sites [61, 60]. By averaging over a large cell population, a map of encounter frequencies summarizes the detected interactions in a large two-dimensional matrix (Fig. 8). There are several refinements, depending on the number of site and resolution. The interaction probability for the interactions between two sites is called 3C, between one site and multiple others (4C) or many sites interact together (5C) and Hi-C, when a low resolution map of all sites is obtained [73, 157, 75]. It is a challenge to reconstruct local physical interactions and the underlying chromatin organization from these averaged maps. Analysis of chromosome capture (CC) data revealed that chromosome distribution in yeast is not uniform. Some chromosomes almost never interact, whereas others show favorable interactions [221]. Hi-C data rely on ligation assays that capture only an instant of the chromatin conformation, which can be a rare event in the large ensemble of chromosomal configuration space. The probabilistic nature of the Hi-C maps reflects both inherent chromatin fluctuation and technical aspects of the methods (cross-linking, digestion and ligation reactions). For that reason, chromatin reconstruction relies on population averaging, which certainly destroys some local organization [200]. However, new techniques of performing single cell CC [189], allow studying the cell-to-cell variability. Features within the chromosome conformation ensemble have recently been found, characterized by increased interactions. These regions are called topologically associating domains (TAD) [197, 91]. As we shall see in this review, modeling ensembles of chromosome conformations has relied on numerical simulations of polymer models [212, 246], with constraints imposed

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

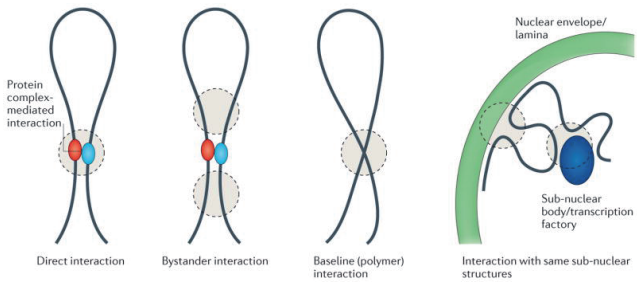


Fig. 7: **Schematic representation of the 3C-based technologies capture loci, located in close proximity.** There are several possibilities for loci to be in close spatial proximity (proteins mediating interaction, Bystander interaction, polymer interaction). All these scenarios are accounted for in the 3C data (reproduced from [60]).

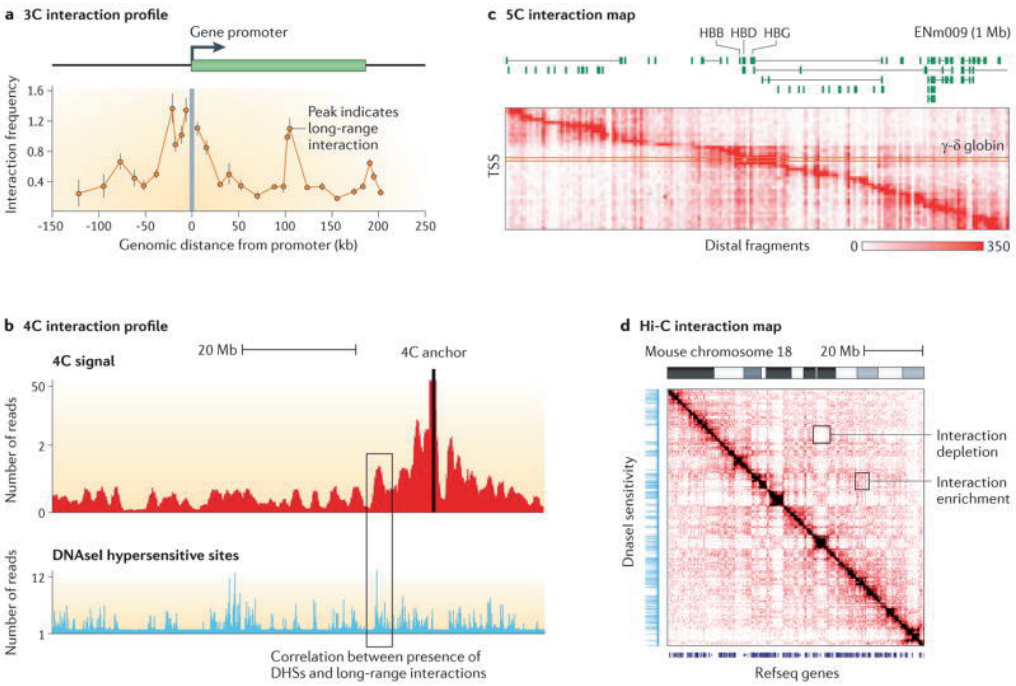


Fig. 8: **3C, 4C, 5C and Hi-C datasets.** A. 3C data B. 4C data. C. An example of a 5C interaction map for the ENm009 region in K562 cells 46. Each row represents an interaction profile of a transcription start site (TSS) across the 1 Mb region on human chromosome 11 that contains the beta-globin locus. D. Hi-C. 3C and 4C data are linear profiles along chromosomes and can be directly compared to other genomic tracks such as DNAseI sensitivity. 5C and Hi-C data are often represented as two-dimensional heat-maps Other genomic features, such as positions of genes or the location of DNAseI hypersensitive sites, can be displayed along the axes for visual analysis of chromosome structural features (reproduced from [60]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

by proximity ligation data [157, 22]. Polymer models are also used to explore the configuration space and DNA unpacking inside the nucleus [200, 89, 272].

2.8 Spatial nuclear organization for gene regulation, DNA maintenance and repair

Spatial nuclear organization is certainly crucial to maintain genome integrity and it plays a key role in double-stranded DNA breaks (DSBs) repair. There are two modes of DSB repair: one is called Non-Homologous-End-Joining, where a break is repaired by a direct relegation [93], the second one is homologous recombination (HR) [160], whereby an unbroken template is used to correctly repair the break (further described in section 9). Interestingly, regions of repair are called foci, that can be visualized using a fluorescent marker - tagging a protein Rad52 involved in DNA repair by facilitating the DNA-strand invasion (see [188, 41] for the peripheral repositioning of the break).

When an artificial break, induced genetically cannot be repaired, the break persists and relocates to the nuclear periphery [188, 201] where it can be anchored. Irreparable breaks that have been processed by the HR pathway associate stably with nuclear pores. At a nuclear pore, a break may be repaired. Possibly the three dimensional configurations of the chromosomes near the nuclear surface play a role in such a process. In summary, DSBs repair involves large-scale motion of the chromosome allowing the break to search for the homologous sequence and for copying the base pairs that have been deleted. During repair, a broken DNA locus scans a large fraction of the nucleus [69, 181]. Interestingly, generating additional damages can affect the dynamics of the homologous-strand, although it is not directly affected. The different steps involved in this repair mechanism remains unexplained. Finally, after a DSB is initiated in the heterochromatin, it moves outside this region to particiapte into the recombination process [49].

There other classical examples where the geometrical organization of the nucleus is involved in regulating its function, including long-range motions in the mammalian nucleus during cell differentiation. In that case, the two X chromosomes needs to meet and to interact. This process can result in X-chromosome inactivation [172]. Large-scale motions are also observed in telomere reorganization [67] and in cancerous cells. Another feature of the geometrical organization is the position of genes: gene regulation depends on their distance to the nucleus surface [266] and in general active genes tend to be found inside the nucleus, as opposed to inactive ones that are located at the periphery.

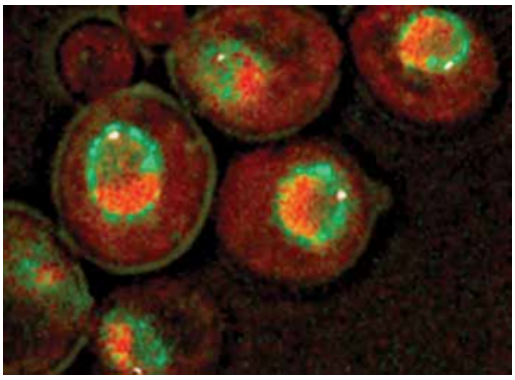


Fig. 9: **Nuclear organization of the yeast.** Visualization of different elements of the yeast nucleus: the nucleolus (red), the yeast spindle pole body (white) and the nuclear envelope (green ring) and a GFP-tagged telomere (brighter green spot). The cells shown here are at different stages of the cell cycle and illustrate the organization within the nucleus (reproduced from [94]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3 Polymer models for modeling DNA dynamics

The renewed interest of classical polymer models [57, 95, 72] answers the need for interpreting, simulating and ultimately understanding the dynamics and structure of the DNA molecule and chromatin. New experiments have called for novel and more accurate polymer models [91] [157]. Basic properties of polymer models have been summarized in [95] and will not be reviewed here and we refer to the classical literature [57, 95] for static and elementary polymer properties, including translocation and reptation [286] through a cylindrical tube.

We recall that DNA is composed of repetitive units of the four fundamental bases guanine (G), adenine (A), thymine (T), or cytosine (C), joint by covalent bonds [282] and can be modeled to the first order as a linear chain composed of connected beads or monomers. A polymer model for a macromolecule is described as a sequence of repetitive structural units. A monomer is a unit representing a sequence of arbitrary length. Once the length is fixed, it defines the scale of the modeling. Thus the same polymer can be used to model a chain with a length ranging from one to hundreds of nanometers. Polymer models have also been used to describe circular DNA (found in prokaryotic cells), where elementary and tractable computations can be performed [134].

How chromatin shall be modeled? are hydrodynamic interactions relevant? For single and double stranded DNA, the local interactions of the chromatin are inferred experimentally from the dynamics of a florescent locus. At a time scale where the chromatin spatial fluctuations is larger than the persistence length L_p, single-stranded DNA ($L_p \sim 2nm$) was suggested to behave as a flexible chain, although the monomer motions are coupled via hydrodynamic interactions [247]. This behavior is well accounted for by the Zimm model [72]. Due to the larger persistence length of double-stranded DNA ($L_p = 50nm$), the inherent molecular rigidity (rod-like behavior) influences the short-time dynamics that are modulated by hydrodynamic effects [113, 205]. At intermediate times, DNA dynamics is quantitatively described by Zimm theory for a semi-flexible chain [113, 114]. More elaborated models take into account bending elasticity to force the alinement of three consecutive monomers, as well as rotational energy of the strand [143].

The contribution of hydrodynamic interactions for describing accurately chromatin dynamics is still under discussion [295], but long-range coherent motion of the chromatin over large distances of the order of several μm is probably synchronized by ATP-dependent motors [35]. At 100ms and more, where independent forces are applied to chromatin, Rouse and generalized type model are used to describe chromatin motion.

In this section, DNA, chromatin and chromosomes are modeled as polymers, discretized into monomers connected by springs. Analysis and stochastic simulations of the models are used to explore the ensemble of conformations. Polymer configurations obtained in a confined domain depend on the confinement itself [72, 8]: the average end-to-end distance of a polymer is smaller in a spherical domain [103] than in a free space [9]. We first review properties of polymer models starting with Rouse (Fig. 10), which is a quite accurate description of chromatin, as confirmed by microscopy experiments [105, 4]. Interestingly, recent CC experiments suggest that the Kuhn length is less than 5kb [231]. This length scale confirms that the bending elasticity may be neglected for modeling chromatin above this small scale. We further discuss more elaborated polymer models, that include interaction energy such as Lennard-Jones or bending elasticity, phenomenological models, based on self-similar properties and characterized by a scaling exponent [183], loop models [29] and the new class of β-polymer model [10] are further discussed for modeling refined chromatin properties. The second part of this section is dedicated to anomalous behavior properties of a monomer.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

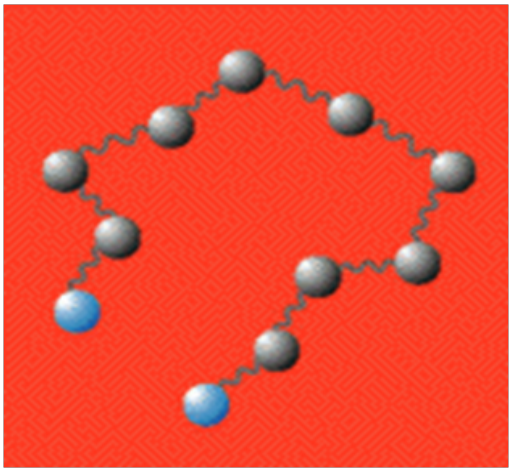


Fig. 10: **Schematic representation of a classical Rouse polymer model, where monomers are connected by identical springs.**

3.1 Elementary polymer chains

Since the mid-20th century, polymer models were built based on a linear chain, where each monomer is connected to only two adjacent monomers. A N-step random walk can be seen as a realization of a polymer of length N, where a monomer is centered at each step point. The result of this random walk is a snapshot of a polymer conformation. When the random particle is not allowed to return to the same point, the realization is interpreted as a self-avoiding polymer (in a good solvent). For such realization, the end-to-end vector ($\boldsymbol{R}_{ee}$) for example, which measures the size and compactness of the polymer is defined by

$$\boldsymbol{R}_{ee} = \sum_{n=1}^{N-1} \boldsymbol{b}_n, \tag{1}$$

where $\boldsymbol{b}_n$ is a step vector of length b (Kuhn length). When the steps are independent, the mean square end-to-end distance is given by [57]

$$\langle \boldsymbol{R}_{ee}^2 \rangle = (N-1)b^2 \tag{2}$$

and for long polymers ($N \gg 1$), the probability distribution function (pdf) of the end-to-end vector $\boldsymbol{R}_{ee}$ was thought to be a Gaussian variable with zero mean and variance Nb^2. A measure of the polymer size is the *radius of gyration* R_g, which accounts for the position of all monomers: it is defined by

$$R_g^2(N) = \frac{1}{N} \sum_{i=1}^{N} \left(\boldsymbol{R_i} - \boldsymbol{R}_{\mathrm{cm}}\right)^2, \tag{3}$$

where $\boldsymbol{R_i}$ is the position of particle i and the center of mass is $\boldsymbol{R}_{\mathrm{cm}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{R_i}$. As the number of monomers N increases (in a free space), the sequences $R_g^2(N)$ converges to the mean value [57]

$$\langle R_g^2 \rangle \approx \frac{Nb^2}{6}. \tag{4}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A *Gaussian chain* model is a continuous polymer model where the distance between two neighboring monomers is a Gaussian variable of zero mean and standard deviation b. The energy of a polymer configuration is given by the sum over the spring energies

$$\phi(\boldsymbol{R}) = \frac{\kappa}{2}\sum_{n=2}^{N}(\boldsymbol{R}_n - \boldsymbol{R}_{n-1})^2, \tag{5}$$

where $\boldsymbol{R} = (\boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_N)$ is the ensemble of monomer positions, connected by a springs of constant $\kappa = dk_BT/b^2$ with standard-deviation b of the distance between adjacent monomers [72], k_B is the Boltzmann coefficient, T the temperature and d the dimensionality (dim 2 or 3). The equilibrium pdf of the polymer is given by the Boltzmann distribution

$$P(\boldsymbol{R}) = \mathcal{N}e^{-\phi(\boldsymbol{R})}, \tag{6}$$

where the normalization coefficient is

$$\mathcal{N}^{-1} = \int_\Omega .. \int_\Omega \prod P(\boldsymbol{R}) d\boldsymbol{R}_1..d\boldsymbol{R}_N = |\Omega| \left[\frac{2\pi}{\kappa}\right]^{(N-1)d/2}, \tag{7}$$

where $|\Omega|$ is the volume of the domain, when it is finite, in which the polymer evolves. The normalization constant is computed over the polymer configuration space, in which case the integral is performed over n-1 variables only. The normalization formula 7 is valid for domains large enough compared to the apparent volume $V_p = \frac{4\pi}{3}R_g^3$, where R_g is the radius of gyration [8].

Another statistical quantity of interest is the distribution of distances between two monomers m and n, which is the sum of Gaussian variables. The mean square separation distance depends on the distances along the chain and using relation 6, it is given by

$$\left\langle (\boldsymbol{R}_n - \boldsymbol{R}_m)^2 \right\rangle = |n-m|\, b^2. \tag{8}$$

3.2 Polymer with self-avoiding interactions in confined domains

In a general polymer model, monomers can interact not only with their nearest neighbors along the chain, but with every one [149, 81]. It is also possible to account for an impenetrable or exclusion volume around monomers, represented by a repulsion force. The associated excluded volume interaction energy ϕ_{EVI} is described by a short-range potential

$$\phi_{EVI}(\boldsymbol{R}_1, ..\boldsymbol{R}_N) = \frac{1}{2}\sigma k_BT \sum_{n,m=1;n\neq m}^{N} \delta(\boldsymbol{R}_n - \boldsymbol{R}_m), \tag{9}$$

where σ is the excluded volume, which is computed from the radius of the monomer and δ is the classical Dirac-function. In one-dimension, excluded volume interactions lead to a stretch polymer, because it cannot overlap anymore. In higher dimensions, the interactions lead to a significant swelling of the polymer size. For ideal polymers, the end-to-end distance follows the scaling $\langle \boldsymbol{R}_{ee}^2 \rangle \approx N$ (formula 8), but with additional exclusion interactions, it can happen that:

$$\langle \boldsymbol{R}_{ee}^2 \rangle \approx N^{2\nu}, \tag{10}$$

where the exponent ν (Flory exponent) depends only on the space dimension: $\nu = 3/4$ for $d = 2$ and $\nu \approx 0.588$ for $d = 3$ [57].

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Elastic and dynamic behaviors of a self-avoiding chain confined inside a cylindrical pore of diameter D_{cyl} have also been studied using a phenomenological Flory approach and molecular-dynamics simulations. This approach is based on minimizing the free energy for the longitudinal end-to-end distance $U(R) = \frac{gR^2}{ND_{cyl}^2} + \frac{D_{cyl}N^2}{g^2R}$, where g is the number of monomers inside a blob ($g \approx (D_{cyl}/a)^{5/3}$) and the excluded volume around each monomer scales like a^3. The chain size was predicted to verify $R \approx Na(\frac{a}{D_{cyl}})^{2/3}$ and the slowest relaxation time of the confined chain in the absence of hydrodynamic effects is $\tau_r \approx N^2a^{2/3}D_{cyl}^{1/3}$ [133], that should be compared with the classical Rouse slowest relaxation mode 22 also proportional to N^2.

The worm-like chain (WLC) model describes a semi-flexible polymers such that the orientation correlation function defined by the tangent vector $t(s) = \frac{dr(s)}{ds}$ decays exponentially along the chain

$$< t(s) \cdot t(0) >= cos(\theta(s)) = e^{-s/P}, \tag{11}$$

where P is the persistence length. We refer to section 3.7 for an elementary construction of the polymer model using the Langevin description. A mean-field approximation model was used in [85] to replace constraints of confinement with a harmonic well approximation (see section 4.6). The simulations of polymers in confined geometries (surface and the interior of a ball cell) allows computing the pressure of DNA packaged in viral capsids. Steady state properties for strongly confined semiflexible polymers, trapped in a closed space or compressed by external forces are further analyzed [229]. Using simulations and scaling argument [42], the escape time of a polymer from a spherical confinement (the polymer has already entered the small hole when it escapes) is predicted to be $\tau \sim \frac{N^{1+\nu}}{\phi^{1/(3\nu-1)}}$, where $\nu = 1/2$ and the monomer concentration is $\phi = \frac{N}{(2R)^3}$ (R is the radius of the cavity). This estimate does not account for the mean time that a polymer end finds the small hole on a cavity, where the polymer will eventually exit.

3.3 The Smoluchowski limit of the Langevin equation for polymer dynamics

We review now polymer models in the context of stochastic processes that involve a dynamics description, which is the basis of analysis and stochastic simulations. We recall that the motion of a particle in a bath of colliding random particles at equilibrium is described by the classical Langevin's equation [237, 219, 238]: for a particle of mass m, the damping force is due friction and is proportional to the particle velocity $\boldsymbol{v}$ only through the friction constant γ. It is assumed to be linear and given by the relation $\boldsymbol{F} = -\gamma\boldsymbol{v}$. The friction constant γ depends on the particle geometry and the viscosity of the fluid. The random and friction forces have a similar origin: in a solvent, the random particles collide with each other and the imbedded particle. When the particle of interest moves in a given direction, more solvent particles are colliding with it from that direction, than from the opposite one, generating the effective friction force.

When the dissipation-fluctuation theorem applies [237], which described the particle moving in a bath of particles at equilibrium, the Newtonian equation of motion [45, 237] becomes

$$m\ddot{\boldsymbol{x}} + \gamma\dot{\boldsymbol{x}} = \sqrt{2\gamma k_B T}\dot{\boldsymbol{w}} - \nabla U(\boldsymbol{x}), \tag{12}$$

where U is a potential energy. In the over-damped limit $\gamma \to \infty$ of the Langevin equation, the damping term is much larger than the inertia term and equation 12 becomes [237]

$$d\boldsymbol{x} = -\frac{1}{\gamma}\nabla U(\boldsymbol{x})dt + \sqrt{2D}d\boldsymbol{w}, \tag{13}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where the diffusion coefficient is $D = k_B T/\gamma$ in dimension 3 [237]. The expression for the diffusion coefficient in dimension two for a cylindrical protein embedded in a membrane depends on the viscosity of that membrane [228]. In this form $\boldsymbol{w}$ is a Gaussian noise with mean zero and variance 1.

3.4 The Rouse model

The Rouse model for a polymer is a collection of beads connected by springs. The monomers are positioned at $\boldsymbol{R}_n$ $(n = 1, 2, ...N)$, following Brownian motion and the spring forces are due to the coupling of the two nearest neighboring monomers. The spring force originates from the potential energy defined in eq.5. In the Rouse model, only neighboring monomers interact [224]. In the Smoluchowski's limit of the Langevin equation 12, the dynamics of monomer $\boldsymbol{R}_n$ driven by the potential $\phi(\boldsymbol{R}_1, .., \boldsymbol{R}_N)$ described by eq.5 and generating the force $-\nabla_{\boldsymbol{R}_n}\phi(\boldsymbol{R}_1, .., \boldsymbol{R}_N)$ is

$$\frac{d\boldsymbol{R_n}}{dt} = -D\kappa(2\boldsymbol{R}_n - \boldsymbol{R}_{n-1} - \boldsymbol{R}_{n+1}) + \sqrt{2D}\frac{d\boldsymbol{w_n}}{dt} \tag{14}$$

for $n = 2, ..N-1$ (the reader should adjust the expression for the first and last monomer). At equilibrium, all beads are centered at zero, but the variance of the distances in a polymer realization is given by

$$< |R_{n+1} - R_n|^2 >= b^2, \tag{15}$$

where b is the standard deviation of the bond length, $\kappa = dk_B T/b^2$ is the spring constant with d the spatial dimension, k_B is the Boltzmann coefficient and T the temperature. A freely-join-chain polymer is a generalized Rouse, where the energy between monomer is given by

$$\phi(\boldsymbol{R}_1, .., \boldsymbol{R}_N) = \frac{\kappa}{2}\sum_{n=1}^{N}\left(|\boldsymbol{R}_n - \boldsymbol{R}_{n-1}| - l_0\right)^2, \tag{16}$$

leading to a steady state configuration, where the mean distance between neighboring monomers is $< |R_{n+1} - R_n| >= l_0$. Taking $l_0 = 0$ is the classical Rouse model. Starting with a given configuration, the relaxation of Rouse polymer to steady state configuration in a free space can be analyzed using the Fourier space (normal or Rouse modes) [72]

$$\boldsymbol{u_p} = \sum_{n=1}^{N}\boldsymbol{R}_n\alpha_p^n, \tag{17}$$

where

$$\alpha_p^n = \begin{cases} \sqrt{\frac{1}{N}}, & p = 0 \\ \sqrt{\frac{2}{N}}\cos\left((n-1/2)\frac{p\pi}{N}\right), & \text{otherwise.} \end{cases} \tag{18}$$

$\boldsymbol{u}_0$ represents the motion of the center of mass and the potential ϕ defined in equation 5 is written with the new coordinate as

$$\phi(\boldsymbol{u_1}, .., \boldsymbol{u_{N-1}}) = \frac{1}{2}\sum_{p=1}^{N-1}\kappa_p\boldsymbol{u}_p^2, \tag{19}$$

where

$$\kappa_p = 4\kappa\sin\left(\frac{p\pi}{2N}\right)^2. \tag{20}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Equations 14 are now decoupled and we obtain a $(N-1)d-$independent Ornstein-Uhlenbeck (OU) processes [277]

$$\frac{d\boldsymbol{u}_p}{dt} = -D_p\kappa_p\boldsymbol{u}_p + \sqrt{2D_p}\frac{d\widetilde{\boldsymbol{w}_p}}{dt}, \tag{21}$$

where each $\widetilde{\boldsymbol{w}_p}$ is an independent d-dimensional Brownian motion, with mean zero and variance 1 and $D_p = D$ for $p = 1..N-1$, while $D_0 = D/N$ and the relaxation times are $\tau_p = 1/D\kappa_p$. The center of mass behaves as a freely diffusing particle. Starting from a straight line, the characteristic time for a Rouse polymer to relax to steady state is constraint by the slowest time constant given by

$$\tau_N = \frac{1}{D\kappa_1} = \frac{1}{4D\kappa\sin\left(\frac{1\pi}{2N}\right)^2} \approx \frac{N^2}{D\kappa\pi^2}. \tag{22}$$

3.5 β-polymer as a generalized Rouse model

What should be modify from the Rouse polymer model so that the the mean-square-displacement (MSD) for a monomer R_c behaves for small time as

$$\langle(\boldsymbol{R}_c(t_0+t) - \boldsymbol{R}_c(t_0))^2\rangle \propto t^\alpha, \tag{23}$$

where $\langle\cdot\rangle$ means averaging over configurations and $\alpha > 0$? is it possible to built a polymer model with a prescribed exponent α? Such construction is possible and the result is a β-polymer [10]. Prescribing the anomalous exponent α imposes intrinsic long-range interactions between monomers, beyond the closest neighbors of the Rouse model.

Starting with the Rouse equations in Fourier's space (eq.21), the coefficients $\tilde{\kappa}_p$ are modified to

$$\tilde{\kappa}_p = 4\kappa\sin^\beta\left(\frac{p\pi}{2N}\right) \tag{24}$$

with a free parameter $\beta > 1$, while $D_0 = D_{\text{cm}}$ and $D_p = D$ for $p > 0$ (see definition of the previous paragraph 3.4). When $\beta = 2$, we recover the Rouse model. The polymer is reconstructed from the inverse matrix transformation eq.18 between the original and the Fourier space. In the procedure, the coefficients α_p^n are not changed, only the exponent in the eigenvalue of equation 24.

This procedure defines a unique ensemble of long-range interactions: modifying the eigenvalues results in long-range monomer-monomer interaction as revealed by computing the potential energy which differs from eq.5. Indeed, using the inverse Fourier transform, we get

$$\tilde{\phi}(\boldsymbol{u}_1, ..\boldsymbol{u}_N) = \frac{1}{2}\sum_p \tilde{\kappa}_p\boldsymbol{u}_p^2. \tag{25}$$

The Rouse transformation in eq. 18 leads to an explicit expression for the interaction between each monomer:

$$\begin{aligned}\tilde{\phi}(\boldsymbol{R}_1, ..., \boldsymbol{R}_N) &= \frac{1}{2}\sum_p \tilde{\kappa}_p\left(\sum_{n=1}^N \boldsymbol{R}_n\alpha_p^n\right)^2 = \frac{1}{2}\sum_{l,m}\boldsymbol{R}_l\boldsymbol{R}_m\sum_p\tilde{\kappa}_p\alpha_p^l\alpha_p^m \\ &= \frac{1}{2}\sum_{l,m}\boldsymbol{R}_l\boldsymbol{R}_m A_{l,m},\end{aligned} \tag{26}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The coefficients are

$$A_{l,m} = \sum_{p=1}^{N-1} \tilde{\kappa}_p \alpha_p^l \alpha_p^m = 4\kappa \frac{2}{N} \sum_{p=1}^{N-1} \sin^\beta \left(\frac{p\pi}{2N}\right) \cos\left(\left(l - \frac{1}{2}\right)\frac{p\pi}{N}\right) \cos\left(\left(m - \frac{1}{2}\right)\frac{p\pi}{N}\right). \tag{27}$$

For $\beta \neq 2$, all monomers are now coupled and the strength of the interaction decays with the distance along the chain, as shown in Fig. 11a-b. For example, the coefficient $A_{50,m}$ between monomer 50 and m depends on the position m (fig. 11b for a polymer of length $N = 100$ and $\beta = 1.5$). The coefficients $A_{n,m}$ obtained for various β are summarized in table 1.

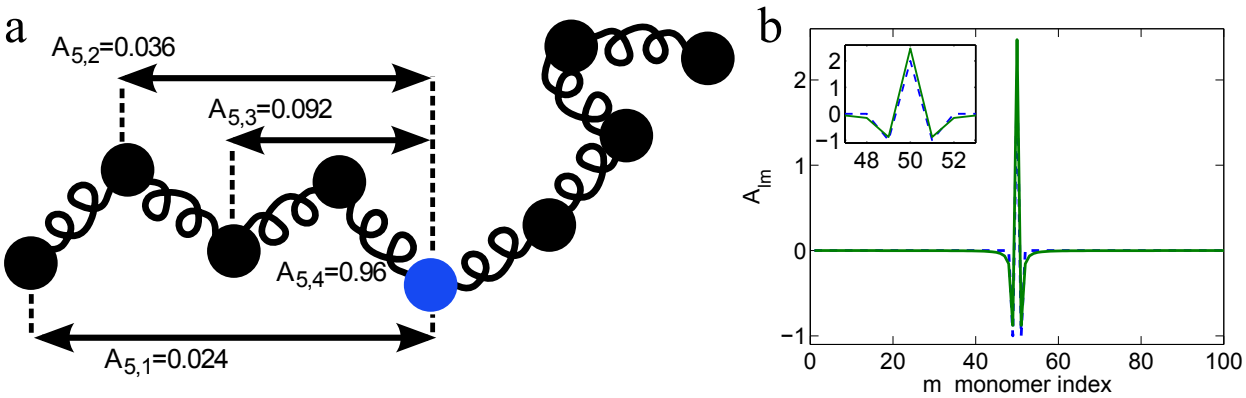


Fig. 11: **The $\beta-$polymer.** (**a**) Representation of a $\beta-$polymer, where all monomers are connected together with a strength that decays with the distance along the chain. The central monomer (blue) interacts with all other monomers in a chain of length $N = 9$ for $\beta = 1.5$ (interaction unit $\kappa = 3/b^2$).(**b**) Monomer-monomer interactions in the modified Rouse polymer model ($\beta-$model). The coefficient A_{lm} (in units of κ) measures the strength of the interaction between two monomers. Shown are the coefficients A_{lm} for the polymer with $\beta = 1.1$, where $l = 50$ and $N = 100$. All monomers interact with each other and the strength of the interaction decays with the distance along the chain (reproduced from [10]).

β	$A_{51,51}$	$A_{51,50}$	$A_{51,49}$	$A_{51,48}$	$A_{51,40}$	$A_{51,30}$
2	2	-1	0	0	0	0
1.5	2.22	-0.95	-0.087	-0.029	-1.07×10^{-3}	-2.2×10^{-3}
1.2	2.40	-0.90	-0.14	-0.054	-3.05×10^{-3}	-7.9×10^{-3}
1	2.55	-0.85	-0.17	-0.073	-5.48×10^{-3}	-16.68×10^{-3}

Table 1: Values of the coefficients $A_{l,m}$ in units of κ for the middle monomer in a polymer chain of length $N = 101$, for different values of β.

The long-range interactions between the monomers in the quadratic potential is linked to pair monomers interacting by

$$\psi(\boldsymbol{R}_1, ..., \boldsymbol{R}_N) = \frac{1}{2} \sum_{ij;\, i>j} a_{i,j} \left(\boldsymbol{R}_i - \boldsymbol{R}_j\right)^2 = \sum_{i=1}^{N} \boldsymbol{R}_i^2 \sum_{j\neq i} \frac{a_{i,j}}{2} - \sum_{ij;\, i>j} a_{ij} \boldsymbol{R}_i \boldsymbol{R}_j. \tag{28}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The coefficients $a_{i,j}$ are related to the $\beta-$polymer $A_{l,m}$, (see eq. 26) by the following formula

$$a_{l,m} = \begin{cases} \sum_{j\neq l} A_{jl}, & l = m \\ -\frac{1}{2}A_{l,m}, & \text{otherwise.} \end{cases} \tag{29}$$

Expression 28 for the potential can be used to reconstruct the amplitude of the interactions from the coefficients of the β-model. In section 3.10.2, we show that the anomalous exponent α (eq. 23) is directly associated with the exponent β of the polymer, by relation $\alpha = 1 - \frac{1}{\beta}$. In summary, within a $\beta-$polymer model, all monomers are coupled and the strength of the interaction decays with the distance along the chain.

3.6 Numerical implementation of polymer simulation using Euler's scheme

The elementary method to simulate stochastic polymer models is based on Euler's scheme [238], where the simulation time step Δt should be chosen such that each monomer moves on average less than the smallest length scale $\Delta x^\star$ of the system. This length $\Delta x^\star$ can be for example the diameter of a small target. In practice, the small spatial scale imposes a constraint on Δt. If we used $\Delta x^\star = f \times \varepsilon$ where f is an extra confident parameter, fixed at 0.2 and ε is a small parameter, then the time step Δt should verify the condition $\sqrt{2D\delta t} \leq \Delta x^\star$. This is usually a serious limitation of the stimulation, that can be relaxed by using an adaptation time step, that can be big far away from the target and refined close enough. In practice, the numerical scheme is

$$\boldsymbol{x}(t+\Delta t) = \boldsymbol{x}(t) + \frac{1}{\gamma}\nabla U^N \Delta t + \boldsymbol{\eta}\sqrt{2D\Delta t}, \tag{30}$$

where U^N is the potential 16. In that case (except the end monomers),

$$\begin{aligned} \boldsymbol{x}_k(t+\Delta t) = \boldsymbol{x}_k(t) - \frac{k}{\gamma}\Bigg(& (\boldsymbol{x}_k(t) - \boldsymbol{x}_{k+1}(t)) - l_0 \frac{\boldsymbol{x}_k(t) - \boldsymbol{x}_{k+1}(t)}{|\boldsymbol{x}_k(t) - \boldsymbol{x}_{k+1}(t)|} \\ & + (\boldsymbol{x}_k(t) - \boldsymbol{x}_{k-1}(t)) - l_0 \frac{\boldsymbol{x}_k(t) - \boldsymbol{x}_{k-1}(t)}{|\boldsymbol{x}_k(t) - \boldsymbol{x}_{k-1}(t)|}\Bigg)\Delta t + \boldsymbol{\eta}\sqrt{2D\Delta t} \end{aligned} \tag{31}$$

where η is a Gaussian variable with zero mean and unit variance. At a reflecting wall, all monomers are reflected according to the classical reflection for Brownian motion, which follows the Snell-Descartes reflection [238] for an isotropic diffusion coefficient, otherwise the co-normal reflection needs to be used [249].

3.7 Implementing excluded volume interactions

Additional interactions such as bending elasticity, which accounts for the persistence length of the polymer and Lennard-Jones forces (LJ) [156], describing self-avoidance of each monomer pairs allow representing more realistic polymers. To account for the LJ-forces, the energy potential is modified to

$$U(\boldsymbol{R}_1, ..\boldsymbol{R}_N) = U_{\text{spring}}(\boldsymbol{R}_1, ..\boldsymbol{R}_N) + U_{LJ}(\boldsymbol{R}_1, ..\boldsymbol{R}_N), \tag{32}$$

where the spring potential is defined by relation 16 and $U_{\text{spring}} = \phi$. The notations are $\boldsymbol{r}_{i,j} = \boldsymbol{R}_i - \boldsymbol{R}_j$, $\kappa = \frac{3}{s_{l_0}^2}$ and s_{l_0} is the standard deviation of the bond length. The Lennard-Jones potential is defined by

$$U_{LJ}(\boldsymbol{R}_1, ..\boldsymbol{R}_N) = \sum_{i,j,i\neq j} U_{\text{LJ}}^{i,j}(\boldsymbol{r}_{i,j}), \tag{33}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and

$$U_{\mathrm{LJ}}^{i,j}(\boldsymbol{r}_{i,j}) = \begin{cases} 4\left[\left(\frac{\sigma}{\boldsymbol{r}_{i,j}}\right)^{12} - 2\left(\frac{\sigma}{\boldsymbol{r}_{i,j}}\right)^{6} + \frac{1}{4}\right] & \text{for } |\boldsymbol{r}_{i,j}| \geq 2^{1/6}\sigma \\ 0 & \text{for } |\boldsymbol{r}_{i,j}| < 2^{1/6}\sigma, \end{cases} \tag{34}$$

where σ is a cut off, for example the size of a monomer. For the chose $l_0 = \sigma$, the springs which materialize bonds cannot cross each other in stochastic simulations of eqs 13 with potential 32.

3.8 Modeling stiffness

To account for some stiffness in polymer model, we can add a bending energy to the total energy of a polymer defined as

$$U_{bend}(\boldsymbol{x}) = \frac{\kappa_{ang}}{2}\sum_{i=1}^{N-1}(\boldsymbol{u}_{i+1} - \boldsymbol{u}_i)^2 = \kappa_{ang}\sum_{i=1}^{N-1}(1 - \boldsymbol{u}_i \cdot \boldsymbol{u}_{i+1}), \tag{35}$$

where $\boldsymbol{u_i} = \frac{\boldsymbol{x}_{i+1}-\boldsymbol{x}_i}{|\boldsymbol{x}_{i+1}-\boldsymbol{x}_i|}$ is the unit vector connecting two consecutive monomers and $\boldsymbol{x}_i$ is the position of the ith monomer. This potential depends on the angle θ_i between two successive monomers with the relation $\boldsymbol{u}_i \cdot \boldsymbol{u}_{i+1} = \cos\theta_i$. The bending rigidity κ_{ang} is related to the persistence length L_p of the polymer by [31]

$$L_p = \frac{\kappa_{ang} l_0}{k_B T}, \tag{36}$$

where k_B is the Boltzmann constant, T is the temperature, l_0 has been defined above. The persistence length quantifies the stiffness of a polymer and can be characterized using the unit tangent vectors $t(s)$ and $t(0)$ at positions s and 0 along the polymer. Averaging over all starting positions, the expectation of the angle ϕ_a, which is the angle between $t(s)$ and $t(0)$, decays exponentially with the distance s along the polymer

$$\langle \hat{u}(s) \cdot \hat{u}(0) \rangle = e^{-s/L_P}. \tag{37}$$

Using the stochastic equation

$$\dot{\boldsymbol{x}} + \nabla U = \sqrt{2D}\dot{\mathbf{w}}, \tag{38}$$

for the total potential

$$U(\boldsymbol{x}) = U^N(\boldsymbol{x}) + U_{bend}(\boldsymbol{x}), \tag{39}$$

$U^N(\boldsymbol{x})$ is the elastic potential defined in Eq.16, the polymer dynamics can be simulated, as shown in figure 12. The parameters are $\kappa_{ang} = 5$ [257] and $L_p = 250nm$. In the coarse-grained simulations of chromatin [8], the two energies U^N and U_{bend} (relation 39) have similar order of magnitude: indeed for the maximum extensibility of DNA about 10% of its total length [37], the length $|r - l_0| \sim 5nm$, and $U_{el} = \frac{1}{2}k|r - l_0|^2 \sim 2 \times 10^{-18}Nm$, while the maximum energy between three consecutive monomers due to bending is $U_{bend} = \kappa_{ang}(1 - \cos\theta) = k_B T L_p/l_0 * 2 \sim 5 \times 10^{-19}Nm$ (for $\cos(\theta) = -1$). Here k_B is the Boltzmann constant and T is the temperature. A snapshot of a stiff polymer is shown in fig. 12 (see also [8]). In comparison, a Rouse polymer would not show bending near the sphere, as the polymer configuration is more compact with a gyration radius proportional to $\sqrt{N}$ (parameters are presented in table 2).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2: Parameters of the simulations

Parameters	**Description**	**Value**
R	Radius of the circular/spherical domain	250 nm
a	radius of the absorbing window	50 nm
l_0	Polymer persistence length	50 nm
D	Diffusion constant	$4 \times 10^{-2}\ \mu\text{m}^2/\text{s}$
γ	Friction coefficient	3.1×10^{-5} Ns/m
k	Spring constant	1.75×10^{-2} Nm^{-1}

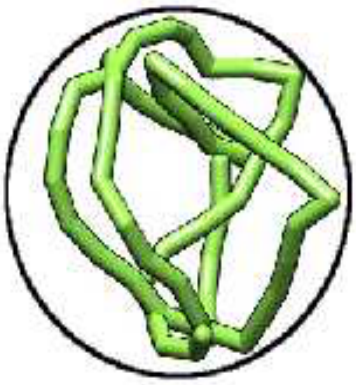


Fig. 12: **Stiff polymer in a ball.** Snapshot of a semi-flexible polymer, locate near the spherical boundary.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.9 Rod-like model

A crude model of polymers is composed of N rigid segments where the angle between two consecutive segments is a random variable θ, uniformly distributed in the interval $[0-2\pi]$. The polymer model consists of a freely jointed chain (see [72] p.8). Each monomer is connected to two neighbors by a rigid bond of length b. This dynamics results in long-range correlations between monomers motion, which is different from Rouse. With this model, it is possible to derive close analytical formula. Indeed this model is tractable and the derived asymptotic formula results can be used to guide intuition for other models [12].

In two dimensions, the rod-polymer model is a collection of $N+1$ monomers (points) at positions $(\boldsymbol{R}_0, \boldsymbol{R}_1, ..., \boldsymbol{R}_N)$ in the complex plane, such that $|\boldsymbol{R}_i - \boldsymbol{R}_{i-1}| = b$ $(i = 1, ...N)$ while $\boldsymbol{R}_0 = 0$. The position at time t of the k-th monomer is

$$\boldsymbol{R}_k(t) = \boldsymbol{R}_{k-1}(t) + b\exp(i\theta_k(t)), \tag{40}$$

where $\theta_k(t) \in [0, 2\pi]$ is the angle between the $x-$axis and the vector $\boldsymbol{R}_k(t) - \boldsymbol{R}_{k-1}(t)$. The position of the k-monomer at time t is the sum of exponential of Brownian motions

$$\boldsymbol{R}_k(t) = b\sum_{j=1}^{k}\exp(i\theta_j(t)). \tag{41}$$

The angles $\theta_1, \theta_2, ..., \theta_N$ are independent Brownian variables on the circle, satisfying the stochastic equation

$$d\theta_j = \sqrt{2D}dB_j, \tag{42}$$

where D is the rotational diffusion coefficient (in s^{-1}), B_i are Brownian variables on a circle of radius 1 with variance 1. The dynamics of the rod-polymer end is represented by the mapping $P(t) = (e^{i\theta_1(t)}, .., e^{i\theta_N(t)})$ on the N-dimensional torus, $\boldsymbol{R}_N(t) = b\Psi(P(t))$ where

$$\Psi : T_N \rightarrow \mathbb{R}^2, \tag{43}$$

$$(\theta_1, .., \theta_N) \rightarrow \sum_{k=1}^{N}\exp(i\theta_k). \tag{44}$$

This approach can be generalized to any dimensions (see [294]).

3.10 Anomalous diffusion in polymer models

Before describing the property of a single monomer, we recall now the general description of anomalous diffusion. The motion of a small molecular probe located at a position $\boldsymbol{R}(t)$ at time t is characterized by the statistics of its second moment time series given for small time $t \ll 1$ by

$$\langle(\boldsymbol{R}(t) - \boldsymbol{R}(0))^2\rangle \sim At^{\alpha}. \tag{45}$$

where $\langle\cdot\rangle$ means averaging over realization and A is a constant. The exponent α characterizes the deviation from normal diffusion: when $\alpha < 1$ (resp. $\alpha > 1$) it is the subdiffusion (resp. superdiffusion) regime. Anomalous diffusion was reported in Richardson study (1926) [216] for turbulence, where the exponent in equation 45 is $\alpha = 3$. Motions for which $\alpha > 2$ are described as ballistic. The dispersive transport of amorphous semiconductors [233] showed that charge carrier exhibits also subdiffusion.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Particles in percolative systems [142] and porous media [140], collective diffusion on solid surfaces [163] and in bulk-surface exchange controlled dynamics in porous glasses [38] also exhibit superdiffusion. Micelles traveling by motor proteins are characterized by an exponent $\alpha = 3/2$, while DNA-binding proteins diffusing along a double stranded DNA can perform subdiffusion [250].

A protein that detaches from a DNA chain and connects to another segment nearby in three dimensions but far along the chain is modeled as a Lévy flights [33, 161]. When a chromatin locus is modeled as a Rouse polymer, the anomalous exponent is $\alpha = 1/2$, but other polymer models lead to other exponents. For the Zimm model, accounting for hydrodynamic interactions, the exponent is $\alpha = 2/3$ [296] while for a reptation-polymer, $\alpha = 1/4$ [57]. Analysis of single particle trajectories of chromatin locus reveals that the anomalous exponent is $\alpha = 0.33$ [104]. Determining the origin of anomalous diffusion in a crowded environment, the cytoplasm or the nucleus remains a challenging problem, because different physical properties (polymer intrinsic mechanical properties, external forces applied to the polymer, etc..) are reflected in that anomalous exponent.

Theoretical models to describe anomalous diffusion are continuous-time random walk (CTRW), obstructed diffusion (OD), fractional Fokker-Planck [138, 139], fractional Brownian motion (FBM), fractional Levy stable motion (FLSM) or fractional Langevin equation (FLE). Phenomenological models [284] based on fractional Langevin equation leads to a MSD that exhibits a power law. The construction of the associated polymer model relies on the Langevin equation with a memory kernel that decays algebraically (see relation 58). This kernel accounts for the properties of the viscoelastic fluid, which slows down any loci dynamics.

We start now by the generalized Brownian motion called *continuous time random walk* (CTRW), a classical model of subdiffusion. In such model, the waiting time between steps and the length of the step are taken from probability distributions $\psi(t)$ and $\lambda(x)$ respectively. Their corresponding mean time and variance are

$$T = \int_0^\infty \psi(t)\, t dt \tag{46}$$

$$\Sigma^2 = \int_{-\infty}^\infty \lambda(x)\, x^2 dx, \tag{47}$$

which can be either finite or not. When the distribution of waiting times is Poissonian and the steps length are Gaussian distributed, it is possible to reproduce in the long-time limit the regular Brownian dynamics. When the waiting time distribution has a long-tailed asymptotic behavior $\psi(t) \approx A_a (\tau/t)^{1+a}$, with $0 < a < 1$, the mean time is infinite. These conditions are used to model a particle trapped for long time (amorphous semiconductors [233]).

For a Gaussian distribution of $\lambda(x)$, the probability distribution $P(x,t)$ of a particle $X(t)$ to be located at position x at time t satisfies the *fractional Diffusion (or Fokker-Planck) equation* (FDE)[141]

$$\frac{\partial P}{\partial t} = {}_0D_t^{1-\gamma} K_\gamma \frac{\partial^2 P}{\partial x^2}. \tag{48}$$

which reduces for $\gamma = 1$ to the classical diffusion equation. ${}_0D_t^{1-\gamma}$ is the Riemann-Liouville operator defined for $0 < \gamma < 1$ by

$${}_0D_t^{1-\gamma} P(x,t) = \frac{1}{\Gamma(\gamma)} \frac{\partial}{\partial t} \int_0^\infty dt' \frac{P(x,t')}{(t-t')^{1-\gamma}}. \tag{49}$$

The fractional differentiation of a power $q \in \mathbb{R}$ is

$${}_0D_t^q t^p = \frac{\Gamma(1+p)}{\Gamma(1+p-q)} t^{p-q}, \tag{50}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where Γ is the Euler's Γ-function. Computing the mean square displacement from equation 48 with initial condition $P(x,0) = \delta(x)$ [141] gives (with $x(0) = 0$)

$$\langle x^2(t)\rangle = \frac{2K_\alpha}{\Gamma(1+\alpha)} t^\alpha. \tag{51}$$

The Lévy flights (LF) model, which is a Markov process with a large jump length distribution is characterized by an asymptotic power law

$$\lambda(x) \approx \frac{\sigma^\mu}{|x|^{1+\mu}}, \tag{52}$$

where the variance is infinite. Trajectories are characterized by an exponent μ [179]. Unlike Gaussian random walk that 'fill' the area ($\mu = 2$), LF-trajectories consist of self-similar clusters, separated by long jumps. The Fourier transform of the function λ is proportional to $|k|^\mu$ (equal to $|k|^2$ for a Gaussian process). The fractional derivative is defined using the Fourier transform [179] as

$$\mathcal{F}\left\{\frac{d^\mu g}{d|x|^\mu}\right\} \equiv -|k|^\mu g(k), \tag{53}$$

where $1 \leq \mu < 2$. Lévy process can be defined though a fractional diffusion equation

$$\frac{\partial}{\partial t} P(x,t) = K^\mu \frac{\partial^\mu}{\partial |x|^\mu} P(x,t), \tag{54}$$

with a generalized diffusion coefficient $K^\mu \equiv \sigma^\mu/\tau$ and $\partial^\mu/\partial|x|^\mu$ is the Riesz-Feller derivative.

The probability density of a LF decays as a power-law $P(x,t) \approx K^\mu t/|x|^\mu$ for $\mu < 2$. Consequently the mean-square-displacement is infinite. A Subdiffusive process $B_H(t)$ can also be generated by the FBM [230, 130]. It has the following properties

$$\begin{aligned} \langle B_H(t)\rangle &= 0, \\ \langle B_H(t)B_H(s)\rangle &= \tfrac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right), \end{aligned} \tag{55}$$

where the parameter H is called the Hurst exponent ($0 < H < 1$). This process is constructed using a generalized Langevin's equation [144] defined as

$$m\frac{dv(t)}{dt} = -\gamma \int_{-\infty}^{t} v(t')K(t-t')dt' + \sqrt{2D}dw^f(t). \tag{56}$$

This equation is not written in the strong damping limit, for which the inertia term remains (m is the mass). The memory kernel $K(t-t')$ relates to the fractional noise $dw^f(t)$, following to the autocorrelation function by the relation

$$\langle dw^f(t)dw^f(t')\rangle = k_B T\gamma K(t-t'). \tag{57}$$

When the kernel is a $\delta-$function, the classical Langevin's equation is recovered. However, for a kernel with a long-time decay (such as a power law), a sub-diffusion behavior is found in the FBM. With the kernel

$$K(t) = 2H(2H-1)\frac{1}{|t|^{-2H+2}}, \tag{58}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the anomalous diffusion exponent is $\alpha = 2H$ and the fractional noise is defined by [145, 63]

$$dw^f(t) = \frac{1}{\Gamma(H+1/2)}\left(\int_0^t (t-t')^{H-1/2}dw(t') + \int_{-\infty}^0 \left[(t-t')^{H-1/2} - (t')^{H-1/2}\right] dw(t')\right). \quad (59)$$

Finally, the variance of the position of a particle governed by eq.56 is

$$\langle |x(t)|^2\rangle \sim \frac{k_B T}{\gamma}\frac{\sin(2H\pi)}{\pi H(1-2H)(2-2H)}t^{2-2H}. \quad (60)$$

To conclude a particle with the characteristic described by eq.56 performs anomalous diffusion with exponent $\alpha = 2-2H$. The FBM is used as a phenomenological model [284] to describe chromatin locus dynamics. The power-law decaying kernel accounts for the motion of the loci in a viscoelastic fluid. In that medium, the motion of a locus as well as the chromosome dynamics are slowed done, resulting in a sub-diffusion regime. Relating the exponent H to the local chromatin properties or the nuclear environment remains a challenge.

3.10.1 Normal and anomalous diffusion regimes for a single monomer

The statistical properties of the kth-monomer motion of a polymer containing N monomers can be computed empirically from its time position $R_k(t)$. Its motion is however non-Markovian, but the motion of the ensemble $(\boldsymbol{R}_1(t),..,\boldsymbol{R}_k(t),..\boldsymbol{R}_N(t))$ is a Markovian process, satisfying the stochastic equations 14. A fundamental characteristic is the MSD (relation 23) for small and intermediate time t. We present here analytical MSD computations for a Rouse and β-polymer. Similar expressions are not known for general polymer model with additional physical forces such as bending elasticity or LJ forces.

3.10.2 Anomalous motion of a Rouse polymer

The position of monomer $\boldsymbol{R}_c$ of a Rouse polymer is given by

$$\boldsymbol{R}_c = \sum_{p=0}^{N-1} \alpha_p^c \boldsymbol{u}_p, \quad (61)$$

where α_p^c are described by relation 18 and the vectors $\boldsymbol{u}_p$ satisfy eqs.21, which form an ensemble of Ornstein-Uhlenbeck processes, for which the variance is computed from each component:

$$\begin{aligned} \sigma_p^2(t) &= \langle |\boldsymbol{u}_p(t) - \boldsymbol{u}_p(0)|^2\rangle = \frac{1}{\kappa_p}(1-\mathrm{e}^{-2t/\tau_p}), \text{ for } p \geq 1\,, \\ \sigma_0^2(t) &= 2D_{\mathrm{cm}}t\,. \end{aligned} \quad (62)$$

The relaxation times are defined by

$$\tau_p = \frac{1}{D\kappa_p},\ p = 1..N-1, \quad (63)$$

while the diffusion constant is $D_{\mathrm{cm}} = D/N$. The shortest time is $\tau_{N-1} \approx 1/(4D\kappa)$ which is half of $\tau_{\mathrm{s}} = 1/(2D\kappa)$ during which a free monomer diffuses a mean square distance between adjacent monomers ($b^2 = 1/\kappa$). The center of mass is characterized by the time scale $\tau_0 \equiv b^2N/D_{\mathrm{cm}} =$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$N^2/(D\kappa)$, that a particle diffuses across the polymer size. For long polymers $\tau_0/\tau_1 \approx \pi^2$. Using relation 62, the MSD of monomer $\boldsymbol{R_c}$ is a sum of independent Ornstein-Uhlenbeck (OU)-variables:

$$\mathrm{var}(\boldsymbol{R_c}(t)) = \langle(\boldsymbol{R}_c(t) - \boldsymbol{R}_c(0))^2\rangle = \frac{d}{\kappa N}\sum_{p=1}^{N-1}\frac{\cos^2\left(\frac{(2c-1)p\pi}{2N}\right)}{\sin^2\left(\frac{p\pi}{2N}\right)}\left(1-e^{-2t/\tau_p}\right) + 2dD_{\mathrm{cm}}t, \tag{64}$$

where d is the spatial dimension. Formula 64 shows the deviation of the MSD compared to Brownian motion. There are three distinct regimes:

1. For short time $t \ll \tau_{N-1}$, $\sigma_p^2(t) \approx Dt$, which is independent of p, (as shown by a Taylor expansion of the exponentials in eq. 62). The sum in eq. 64 gives

$$\mathrm{var}(\boldsymbol{R_c}) \approx 2dDt, \tag{65}$$

   which characterizes diffusion.

2. For large time $t \gg \tau_1$, after relaxation, the exponential terms in relation 64 becomes independent of t. Only the first term in Eq. 62 corresponding to the diffusion of the center of mass gives the time-dependent behavior. This regime is dominated by normal diffusion, with diffusion coefficient D/N.

3. For intermediate times $\tau_{N-1} \ll t \ll \tau_1$, such that $2t/\tau_p > 1$, the sum of exponentials contributes to eq.64. The variance 64 can be approximated by

$$\mathrm{var}(R_c) \approx 2\int_{p_{\min}}^{N-1}\frac{\cos^2\left(\frac{(2c-1)p\pi}{2N}\right)}{\sin^2\left(\frac{p\pi}{2N}\right)}dp, \tag{66}$$

   where $p_{\min}$ is such that $\tau_{p_{min}} = 2t$. We shall see below that $\mathrm{var}(\boldsymbol{R}_c) \sim t^{1/2}$. Due to the behavior of a Rouse polymer, all internal relaxation times τ_k contribute to the intermediate time, leading to anomalous diffusion for a Rouse monomer. The time interval can be arbitrarily long with the size N of the polymer.

We present below the analytical derivation of the power law behavior.

3.10.3 Monomer motion characterized by a power-law decay time

To describe the large range of anomalous exponent of chromatin locus diffusion, the Rouse potential ϕ_{Rouse} (relation 5) is introduced in eq. 56 [284], so that every monomer is characterized by a subdiffusion regime. The motion of a monomer is described by

$$\gamma\int_0^t \frac{d\boldsymbol{R}_c(t')}{dt}K(t-t')dt' = \sqrt{2D}dw^f(t) + \nabla\phi_{\mathrm{Rouse}}(\boldsymbol{R}_1, ...\boldsymbol{R}_N). \tag{67}$$

The power-law kernel K, described in 58, accounts for the motion of the loci in a viscoelastic fluid, which slows done the chromosome dynamics, leading to a sub-diffusion regime. For the middle monomer, the long-time asymptotic of the MSD is given by [284]

$$\langle(\boldsymbol{R}_c(t) - \boldsymbol{R}_c(0))^2\rangle = \frac{3k_BT}{N\gamma}\frac{\sin(\alpha\pi)}{\pi\left(1-\frac{\alpha}{2}\right)(1-\alpha)\alpha}t^\alpha + \sum_{p=1}^{\infty}\frac{12k_BT}{k_{2p}}\left[1 - E_{\alpha,1}\left(\frac{k_{2p}}{N\gamma\Gamma(3-\alpha)}t^\alpha\right)\right], \tag{68}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where $E_{\alpha,1}$ is the generalized Mittag-Leffler function $E_{\alpha,\beta}(z) = \sum_{j=0}^{\infty} \frac{z^j}{\Gamma(\beta+\alpha j)}$, $k_{2p} = [3\pi^2 k_B T/(Nb^2)](2p)^2$. Here, for short times, the MSD of the middle monomer is

$$\langle (\boldsymbol{R}_c(t) - \boldsymbol{R}_c(0))^2 \rangle \sim t^{\alpha/2}, \tag{69}$$

where $0 < \alpha < 1$. At long times, the motion is governed by that the center of mass leading to

$$\langle (\boldsymbol{R}_c(t) - \boldsymbol{R}_c(0))^2 \rangle \sim t^{\alpha}. \tag{70}$$

3.10.4 Correlation function for a Rouse and $\beta-$polymer

We have already introduced the β-polymer model in section 3.5. This class of polymer resolves the following question: construct a generic polymer model for which the anomalous exponent of a locus is prescribed. A large variability of the anomalous exponent for the dynamics of a chromatin locus cannot be accounted for by the viscoelasticity alone, subdiffusion with an exponent $\alpha > 0.5$ can also appear in some polymer model [11], a case that involves a deterministic force or directed motion or other monomer interactions (self-avoiding and bending interactions).

We present here the computation of the cross-correlation function for monomer $\boldsymbol{R}_c$ using eq. 61. For $N \gg 1$ and intermediate times $\tau_{N-1} \leq t \leq \tau_1$, we show that

$$\langle (\boldsymbol{R}_c(t_0+t) - \boldsymbol{R}_c(t_0))^2 \rangle \propto t^{1-\frac{1}{\beta}}. \tag{71}$$

Using 25, for any time t_0 and t, a direct computation from eq. 61, with $\boldsymbol{R}_c = \sum_{p=0}^{N-1} \alpha_p^c \boldsymbol{u}_p$ and given the correlation function of $\boldsymbol{u}_p$ which is $\langle |\boldsymbol{u}_p(t_0+t) - \boldsymbol{u}_p(t_0)|^2 \rangle = \frac{2}{\tilde{\kappa}_p}(1 - e^{-t/\tau_p})$, we get

$$\langle (\boldsymbol{R}_c(t_0+t) - \boldsymbol{R}_c(t_0))^2 \rangle = \frac{d}{\kappa N} \sum_{p=1}^{N-1} \frac{\cos^2\left(\frac{(2c-1)p\pi}{2N}\right)}{\sin^{\beta}\left(\frac{p\pi}{2N}\right)} \left(1 - e^{-t/\tau_p}\right) + 2dD_{\mathrm{cm}}t, \tag{72}$$

where $\tau_p = 1/D\tilde{\kappa}_p$. We plot in fig.13b the cross-correlation function of a monomer computed from Brownian simulations of a polymer with length $N = 128$ and compare the cross-correlation function for $\beta = 3/2$ with the regular Rouse model ($\beta = 2$). In the intermediate time regime, where $\tau_{N-1} \ll t \ll \tau_1 \approx N^{\beta}\tau_{N-1}$ (anomalous regime), eq.72 can be estimated. For large N, using Euler-Maclaurin formula, the cross-correlation function is approximated for $1 \leq \beta \leq 2$ by [10]

$$\begin{aligned} \langle (\boldsymbol{R}_c(t_0+t) - \boldsymbol{R}_c(t_0))^2 \rangle \approx & \frac{d}{\kappa N}\left(\frac{2N}{\pi}\right)^{\beta}\left[\frac{1}{\beta-1} + \left(\frac{1}{2} + \frac{\beta}{12}\right)\left(1 - e^{-t/\tau_1}\right)\right. \\ & \left. + \frac{\pi}{\Gamma(2-1/\beta)\sin(\pi/\beta)\beta}\left(\frac{t}{\tau_1}\right)^{1-\frac{1}{\beta}}\right] + \frac{2dDt}{N}. \end{aligned} \tag{73}$$

Note that this formula uses that $-\Gamma(1/\beta - 1) = \frac{\pi}{\Gamma(2-1/\beta)\sin(\pi/\beta)} > 0$ and $\Gamma(2 - 1/\beta) > 0$, because $\beta \geq 1 > 1/2$. We conclude that for intermediate times ($\tau_{N-1} \leq t \leq \tau_1$), the center of mass diffusion does not contribute to the process and thus the cross-correlation function scales with time as shown in eq.71: while the exponent α is an indicator of anomalous dynamics, it is a challenging task to relate it to the hidden underlying driving forces, local geometrical organization and crowding organization.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

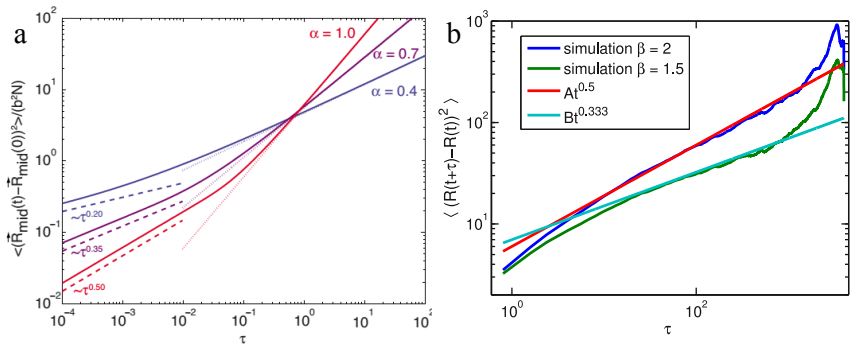


Fig. 13: **Anomalous diffusion properties of a single monomer.** (**a**) MSD computed for the midpoint monomer $\langle(\boldsymbol{R}_{\rm mid}(t)-\boldsymbol{R}_{\rm mid}(0))^2\rangle/b^2N$ of a fractional Langevin motion for a polymer versus the dimensionless time $\tau = t/\left[N^2b^2\gamma/k_BT\right]^{1/\alpha}$ for $\alpha = 1.0$ (red), $\alpha = 0.7$ (purple), and $\alpha = 0.4$ (blue) see [284]. Analytical solution of (eq.68) (solid lines) are represented for the three values with the long-time asymptotic behavior (dotted) and short-time scaling of $\tau^{\alpha/2}$ (dashed). (**b**) Brownian simulations of a polymer (length $N = 128$) for different values of β. The correlation function is computed from simulated trajectories. For $\beta = 3/2$ (green line), the anomalous exponent is $\alpha = 0.33$ at intermediate times (eq.71). The trend line (cyan) is also plotted. For the Rouse model $\beta = 2$ (blue line), the anomalous exponent is $\alpha = 0.5$. Also plotted is the trend line (red) according to eq.71.

4 Polymer looping and the search for a small target: a mean passage time study

This section is dedicated to polymer looping and the estimation for the rate of this process. We start with early models [202], followed by non-Markovian dynamics [287, 288, 261, 64, 254, 99, 100] and then asymptotic approaches [13, 9, 14]. We also present simulation results for the search for a small target and the associated Mean First Encounter Time (MFET), as asymptotic computations are still lacking. For a recent review of the search process for a small target by Brownian particles, we refer to the Narrow Escape Theory [119, 118, 55]. We recall now the context of chromatin looping.

DNA short and long-range loops are observed in chromatin at various times and spatial scales (as we will discussed in the next section 5). A loop is formed when one piece of the chromatin is brought into close proximity of another one, although the two parts could be located far away in terms of the genomic distance. The continuous change of chromatin configuration that lead to looping can result in gene direct interaction [209] and expression [148], modulation, activation or repression. DNA and polymer looping have been investigated experimentally [66, 76], numerically [47] and analytically [288, 261, 202, 13, 9] over the past 40 years in the context of polymer physics with recent renewed interest in the recent years. Some early computations for the time a polymer knot become untied [286] are not discussed much here, but these computations are quite relevant to interpret data about two-tagged locus located on the same DNA.

The questions to consider are: estimating the mean looping time in free and confinement domain as well as the search time for a polymer locus to find a small target, which tends to infinity as the size of the target tends to zero. Deriving analytical formula allows exploring continuously the parameter space (where variables are the target size, the polymer length, the bond strength, etc...) at low cost compared to numerical simulations. In the context of the nucleus, such formulas provide estimates for the rate that two loci come into close proximity or the time a DNA frag-

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ment meets a small target driven by random collision. They are used to interpret data (see section 5).

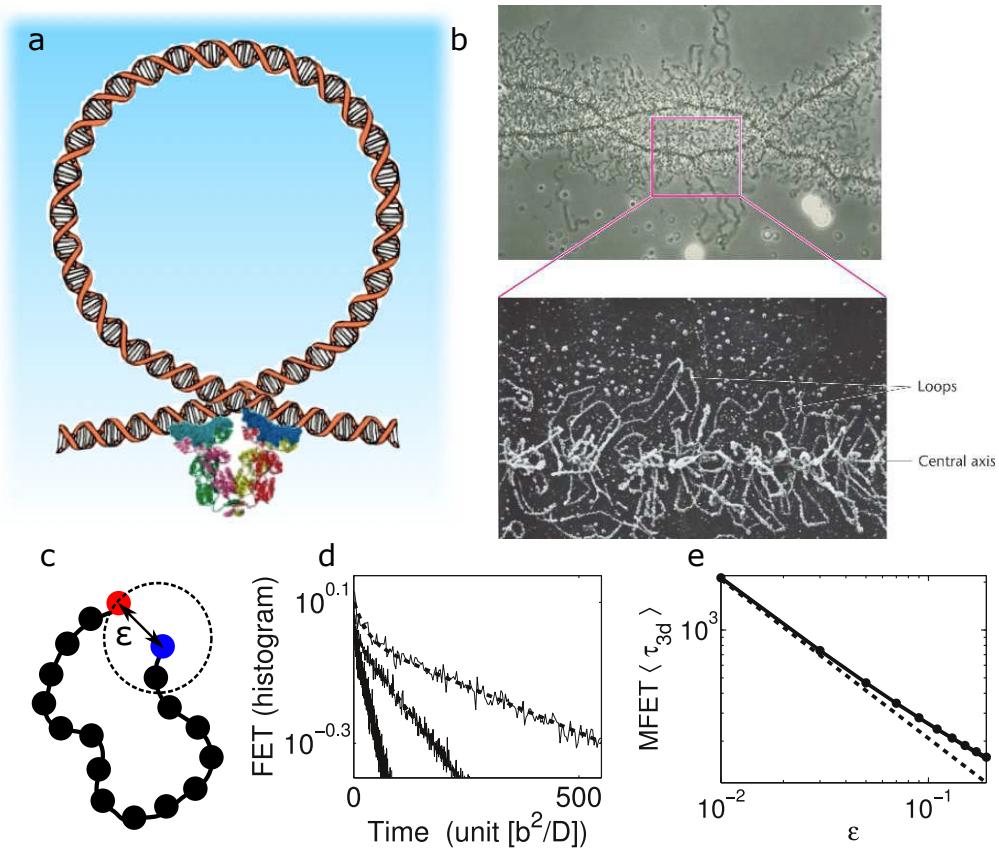


Fig. 14: **DNA looping**: **(a)** Schematics of a DNA loop **(b)** Image of loops along the chromatin **(c)** Scheme of a polymer loop, occurring when the two ends are located at a distance ε from one another. **(d)**) Histogram of the first encounter times (FET) obtained from Brownian simulations in three dimensions (full line) and fitting with two exponentials (dashed line) for different polymer length ($N = 16, 32, 64$ (left to right) and $\varepsilon = 0.1b$. The distribution of looping time is well approximated by a single exponential for small N, but two are necessary for larger N. **(e)** MFET as a function of the radius ε in three dimensions. Comparison of the Brownian simulations (full line) with the reciprocal of the first term in the expansion of the first eigenvalue (eq.114)(dashed line) and eq.175) (circles).

The mean time for two polymer ends to meet (Fig. 14), starting from any given configuration has several implications for DNA looping because a gene can be activated when a transcription factor bound far away from the promoter site is brought near the active site [66, 209, 76]. The mean first encounter time (MFET) is defined as the first arrival time for the end monomer into a ball of radius ε, centered at the other polymer end (Fig. 14c). The MFET depends on the radius ε and the polymer length N (measured in the number of monomers). For a Rouse polymer, characterized by the relaxation time τ_p [72] where the slowest is proportional to N^2 ([254] and section 3.4), the MFET depends both on the initial end-to-end distribution [261, 287, 82] and the radius ε of a ball located around one end (Fig. 14c). These two time scales are reflected in the asymptotic regimes, identified numerically [47, 78] and characterized by the ratio $\sqrt{N}\varepsilon/b$ (b is the standard deviation of the bond

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

length). When this ratio is of order 1, the MFET depends on ε and scales as $N^{3/2}$, when it is $\gg 1$, it is dominated by N^2 and is independent of ε. In general, the MFET shows both scales [273] in N as we shall see here.

For a worm like-chain model [125] with hydrodynamic forces, self-avoidance and coulomb forces, the MFET depends on several parameters such as the polymer length and the bending elasticity, as shown numerically [206, 256]. The MFPT has been estimated using analytical methods [288, 261, 202, 273, 82], numerically [47] and experimentally [218, 7] and also when one end was tethered to a surface [271].

To derive the dependency of the MFET on N, we shall recall its formulation as a mean first passage time to a small boundary, [237, 202] which is a boundary value problem for the second order Laplace equation in a high-dimensional space, defined by the polymer configuration space. The MFET is computed by expanding the first eigenvalue for the Fokker-Planck operator associated to the stochastic Rouse dynamics. Although the Markovian aspect of looping and search process for a small target by a polymer locus was established in early models [202], the non-Markovian aspect was recently considered [254, 99, 100], but the search time is in fact well-approximated as a Poissonian process [8, 9, 14].

4.1 Brief analytical formulation

The first looping time between two monomers is the First Encounter Time τ_e for two monomers n_a, n_b to come into a distance $\varepsilon < b$. It is defined by

$$\tau_e = \inf\{t > 0 \text{ such that } |\boldsymbol{R}_{n_a}(t) - \boldsymbol{R}_{n_b}(t)| \leq \varepsilon\}, \tag{74}$$

where $\boldsymbol{R}_{n_a}$ and $\boldsymbol{R}_{n_b}$ follows for example the Rouse equation 14. The goal of the section is to present several analytical studies to compute the Mean First Encounter Time (MFET) $\langle \tau_e \rangle$.

4.2 Szabo, Schulten& Schulten approach and later analysis

Applying the first passage times theory [236, 237, 238] to diffusion controlled reactions, in a pioneering article Szabo, Schulten& Schulten [261] studied intramolecular reactions of polymer end groups by using differential equations. Depending on the boundary conditions, chosen to be absorbing, or radiative boundary condition (partial reflection), they derived various asymptotic expressions for the chemical reaction time. For a diffusion process $X(t)$ satisfying the stochastic equation

$$\dot{\boldsymbol{X}} = \boldsymbol{a}(\boldsymbol{X}) + \boldsymbol{b}(\boldsymbol{X})\dot{\boldsymbol{w}} \tag{75}$$

where $\boldsymbol{a}$ is a deterministic drift, $\boldsymbol{b}$ the diffusion tensor and $\boldsymbol{w}$ the classical Wiener δ-correlated noise, the MFPT for the arrival time $\mathbb{E}\tau(\boldsymbol{x})$ starting at position $\boldsymbol{x} = \boldsymbol{X}_0$ is solution of

$$L^*_{\boldsymbol{x}}\left(\mathbb{E}\tau(\boldsymbol{x})\right) = \quad \sum_{i,j=1}^{d} \sigma^{i,j}\left(\boldsymbol{x}\right) \frac{\partial^2 \mathbb{E}\tau(\boldsymbol{x})}{\partial x^i \partial x^j} + \sum_{i=1}^{d} \boldsymbol{a}^i\left(\boldsymbol{x}\right) \frac{\partial \mathbb{E}\tau(\boldsymbol{x})}{\partial x^i} = -1, \tag{76}$$

in dimension d, where $\boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x})\boldsymbol{b}^T(\boldsymbol{x})$. An absorbing boundary condition is imposed on the target and reflecting boundary condition are given on other part of the boundary. In particular, in spherical coordinate in d-dimensions, for a gradient field $\boldsymbol{a} = -D\beta\nabla U$, equation 76 becomes

$$\frac{1}{r^{d-1}} \frac{d}{dr}\left(r^{d-1} D(r) \frac{d}{dr} \mathbb{E}\tau(r)\right) - D\beta \frac{dU(r)}{dr} \frac{d\mathbb{E}\tau(r)}{dr} = -1 \tag{77}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with reflecting boundary condition at $r = R$ and absorbing at $r = a$. After a double integration, the mean time is

$$\mathbb{E}\tau(r) = \int_a^r \frac{x^{1-d}}{D(y)} \exp(\beta U(x))dx \left(\int_x^R \frac{y^{d-1}}{D(x)} \exp(-\beta U(y))dy \right) \tag{78}$$

and the mean $\mathbb{E}\tau(r)$ averaged over the steady state distribution $p_e(x) = \frac{x^{d-1}\exp(-\beta U(x))}{\int_a^R y^{d-1}\exp(-\beta U(y))dy}$ is

$$\langle \mathbb{E}\tau \rangle = \frac{1}{\int_a^R x^{1-d}\exp(-\beta U(x))dx} \int_a^R y^{1-d}\exp(\beta U(x))dx \left(\int_x^R \frac{y^{d-1}}{D(x)} \exp(-\beta U(y))dy \right)^2 \tag{79}$$

The dynamics of the end-to-end distance $r = |R_N - R_0|$ is described by a Smoluchowski-type equation and the overall effect of the polymer can be approximated by harmonic spring potential, i.e., $U(r) = \frac{d}{2\beta L^2} r^2$, where L is the mean distance between the polymer ends. When the radius R is large, but a is small, the asymptotic expansion of 79 for $d = 3$ is

$$\frac{3D}{L^2}\langle \mathbb{E}\tau \rangle = \frac{\sqrt{\pi}}{2\chi} + (ln2 - 1) - \chi\sqrt{\pi}2 + \frac{4}{3}\chi^2 + \mathcal{O}(\chi^2), \tag{80}$$

where $\chi = (3/2)^{1/2}\frac{a}{L}$. In addition, numerical simulations suggest that the passage time distribution is well approximated by a single exponential, where the rate is $\frac{3D}{L^2}\langle \mathbb{E}\tau \rangle$. This approach was later on improved in [202], followed by a discussion of the validity of the asymptotic expansion, depending whether the radius a is larger or smaller compared to b (equilibrium mean length of a single bond) and $\langle (R_N^2 - R_0)^2 \rangle_{eq} = L^2 = Nb^2$. Indeed, the equilibrium distribution $\rho(r)$ of the end-to-end distance r can be approximated by a Gaussian with the mean squared distance L^2,

$$\rho_{eq}(r) = \left(\frac{3}{2\pi L^2} \right)^{3/2} \exp\left(-\frac{3r^2}{2L^2} \right), \tag{81}$$

the following radial Laplace equation is solved with a weight ρ_{eq} and an absorbing boundary conditions at $r = a$:

$$\begin{aligned} D\frac{1}{\rho_{eq}(r)r^2}\frac{\partial}{\partial r}\left(r^2 \rho_{eq}(r)\frac{\partial}{\partial r} \right)\mathbb{E}\tau(r) &= -1, \\ \mathbb{E}\tau(a) &= 0. \end{aligned} \tag{82}$$

The effective diffusion coefficient D was chosen empirically, based on numerical integration $D = 2D_0$, (D_0 is the diffusion coefficient of a single monomer in the Rouse equation) to account for the diffusion of the two ends, neglecting the effect of all other monomers. Equation 81 assumes that the local equilibrium approximation is valid, but in reality, it is valid only when the rate of approach to local equilibrium is much faster than the rate of the end monomers to meet. The time scale to equilibrium is governed by the largest Rouse mode relaxation time (see eq. 22) $\tau_R \approx \frac{L^2}{\pi^2}$ in units ($b = D = 1$) and does not depend on the radius a. The local equilibrium condition can be satisfied only when the contact distance a is very small. In the small a limit, the mean first passage time, predicted by eq. 82 is

$$\langle \mathbb{E}\tau \rangle \approx \left(\frac{2\pi}{3} \right)^{1/2} \frac{N^{3/2}}{12D_0 a}. \tag{83}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

If a is much smaller than $b/\sqrt{N}$, the rate to equilibrium is faster than trapping, for which formula 83 is valid.

In summary, in the limit of small a for fixed N, $\langle \mathbb{E}\tau \rangle$ is of the order of $N^{3/2}/a$, while in the limit of large N at fixed a, $\langle \mathbb{E}\tau \rangle$ is of order N^2 and independent of a. But as we shall see below, in general, the looping time formula combines both behavior in $N^{3/2}/a$ and N^2. A general computation of the asymptotic behavior for the looping time requires considering the motion of the entire polymer and not simply the two end distance, as we shall see below [13].

4.3 Closed loop configuration space for a rod-like polymer

We introduce in sub-section 3.9 rod-like polymer models. We describe now the configuration space of loops. The MFET $\langle \tau_\epsilon \rangle$ for the two ends of a rod-polymer is equivalent to the one for a Brownian particle $P(t) = (e^{i\theta_1(t)}, .., e^{i\theta_N(t)})$ located on the torus T_N in dimension N (length of the polymer) to meet for the first time the boundary of the domain

$$B_\epsilon = \{(\theta_1, .., \theta_N) \in T_N \text{ such that } \left| \sum_{k=1}^{N} \exp(i\theta_k) \right|^2 \leq \epsilon^2\}. \tag{84}$$

The domain B_ϵ is a tubular neighborhood of a $N-2-$dimensional manifold defined by the B_0 (taking $\epsilon = 0$ in equation 84). B_0 is of codimension 2, because it is the set of points defined by two independent equations

$$\sum_{k=1}^{N} \sin\theta_k = 0 \text{ and } \sum_{k=1}^{N} \cos\theta_k = 0. \tag{85}$$

The boundary ∂B_ϵ is represented for $N = 3$ and $N = 4$ in Fig.16 for $\epsilon = 0.2b$. The boundary ∂B_ϵ is analyzed by setting $\epsilon = 0$ corresponding to closed configurations (Fig.15): for $N = 3$, the ensemble of closed chains (Fig. 15B) consists of two equilateral triangles (for each orientation) with one vertex at zero. This ensemble is invariant by planar rotations centered at zero. Under this action, the ensemble consists of two orbits and any reflection (preserving zero) permutes these two orbits. For $N = 4$, the three different configurations are shown in Fig. 15C): it consists of a polygon and two non-isomorphic configurations. These structures are connected by continuous deformations. In that case, the boundary ∂B_ϵ is made of domains, connected by narrow structures (Fig. 16b). The description of B_0 for large N is quite difficult due to lack of parametrization [12].

4.4 Looping for a Rouse polymer

The two ends $\boldsymbol{R}_N, \boldsymbol{R}_1$ meets when their distance is less than $\varepsilon < b$, that is

$$|\boldsymbol{R}_N - \boldsymbol{R}_1| \leq \varepsilon \tag{86}$$

In Rouse coordinates, $\boldsymbol{u_p} = \sum_{n=1}^{N} \alpha_p^n \boldsymbol{R}_n$ [72] where α_p^n are defined in 18, condition 86 is

$$\left| 2\sqrt{\frac{2}{N}} \sum_{p \text{ odd}} \boldsymbol{u}_p \cos(p\pi/2N) \right| \leq \varepsilon. \tag{87}$$

The end-to-end encounter is independent of the center of mass (coordinate $\boldsymbol{u}_0$ 17). The MFET is the mean first passage time of the $(N-1)d$-dimensional stochastic process

$$\boldsymbol{u}(t) = (\boldsymbol{u}_1(t), ..\boldsymbol{u}_{N-1}(t)) \in \Omega \times \Omega ... \times \Omega = \tilde{\Omega}, \tag{88}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

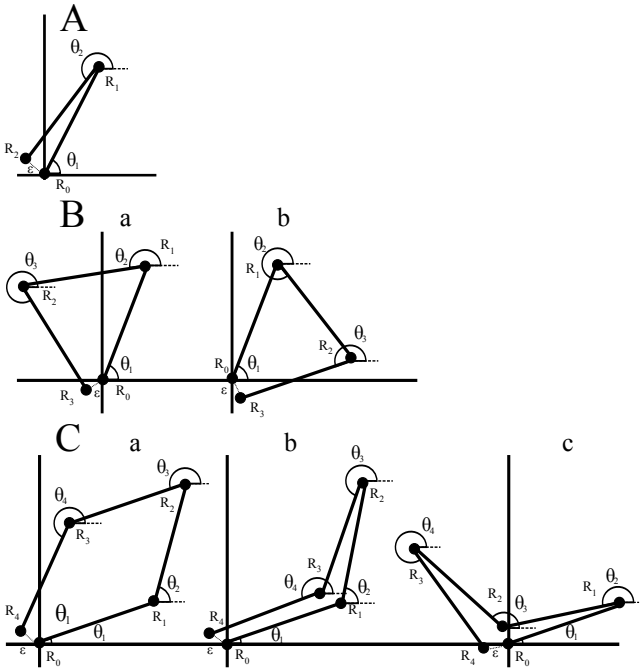


Fig. 15: **Schematic representation of polymer loops for different lengths**. Monomers positions are $\boldsymbol{R}_k$ while $\boldsymbol{R}_0$ is fixed at zero. θ_k is the angle between $\boldsymbol{R}_k - \boldsymbol{R}_{k-1}$ and the positive direction of the x axis. ϵ is the absorption distance. **A**: two bonds polymer ($N = 2$): there is only one closed configuration (up to rotation). **B**: three bonds polymer ($N = 3$), showing closed that the only configuration is an equilateral triangle (the loop can be either clock-wise or counter clock-wise). **C**: four bonds polymer ($N = 4$), showing three possible closed configurations [12].

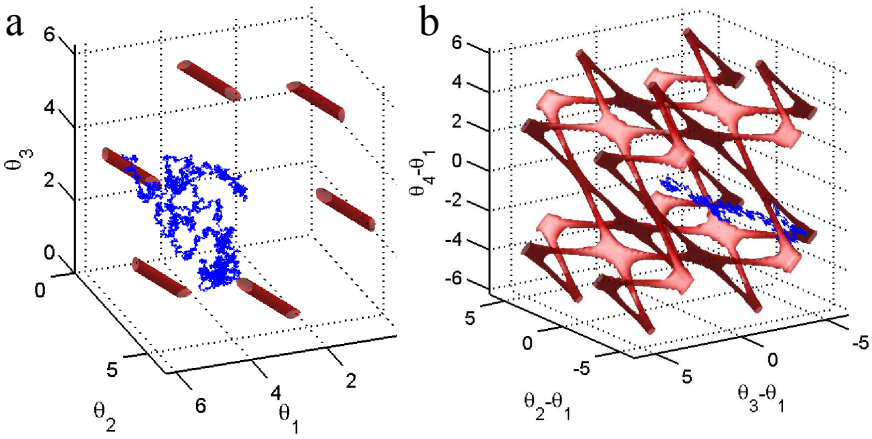


Fig. 16: **The boundary of B_ϵ for a rod-polymer. a)**: for $N = 3$ and $\epsilon = 0.2$ in $[0, 2\pi] * [0, 2\pi] * [0, 2\pi]$. Although figure (a) show 6 disconnected, there are only two of them in a unitary torus. A Brownian trajectory (blue) in the angle space moves until it is absorbed.**b)**: for $N = 4$ and $\epsilon = 0.2b$. The manifold consists of domains connected by narrow tube, corresponding to deformation of loops [12].

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where $\Omega = \mathbb{R}^2$ or $\mathbb{R}^3$ and $\boldsymbol{u}_p$ satisfies the OU-equations 21 to the boundary of the domain

$$S_\epsilon = \{P \in \tilde{\Omega} \text{ such that } dist(P, \mathcal{S}) \leq \frac{\varepsilon}{\sqrt{2}}\}, \tag{89}$$

where $dist$ is the Euclidean distance and

$$\mathcal{S} = \{(\boldsymbol{u}_1, ..\boldsymbol{u}_{N-1}) \in \tilde{\Omega} \big| \sum_{p \text{ odd}} \boldsymbol{u}_p \cos(p\pi/2N) = 0\} \tag{90}$$

is a submanifold of codimension d in $\tilde{\Omega}$. The probability density function (pdf) $p(\boldsymbol{u}(t) = \boldsymbol{x}, t)$ satisfies the forward Fokker-Planck equation (FPE) [237]

$$\begin{aligned} \frac{1}{D}\frac{\partial p(\boldsymbol{x}, t)}{\partial t} &= \Delta p(\boldsymbol{x}, t) + \nabla \cdot (\nabla \phi \, p(\boldsymbol{x}, t)) = \mathcal{L}p, \\ p(\boldsymbol{x}, 0) &= p_0(\boldsymbol{x}), \end{aligned} \tag{91}$$

with boundary condition $p(\boldsymbol{x}, t) = 0$ for $x \in \partial S_\epsilon$, $p_0(\boldsymbol{x})$ is the initial distribution and the potential $\phi(u_1, .., u_N) = \frac{1}{2}\sum_p \kappa_p \boldsymbol{u}_p^2$ was introduced in 19.

The solution of equation 91 can be expanded in eigenfunctions

$$p(\boldsymbol{x}, t) = \sum_{i=0}^{\infty} a_i w_{\lambda_i^\epsilon}(\boldsymbol{x}) e^{-\lambda_i^\epsilon t D} e^{-\phi(\boldsymbol{x})}, \tag{92}$$

where a_i are coefficients, $w_{\lambda_i^\epsilon}(\boldsymbol{x})$ and λ_i^ϵ are the eigenfunctions and eigenvalues respectively of the operator $\mathcal{L}$ in the domain $\Omega_\epsilon = \tilde{\Omega} - S_\epsilon$. The probability distribution that the two ends have not met before time t is the survival probability

$$p(t) = Pr\{\tau_\epsilon > t\} = \int_{\Omega_\epsilon} p(\boldsymbol{x}, t) dx, \tag{93}$$

and the first looping time is $\tau_\epsilon = \inf\{t > 0, \boldsymbol{u}(t) \in \partial S_\epsilon\}$. Using expansion 92, $p(t) = 4\sum_{i=0}^{\infty} C_i e^{-\lambda_i^\epsilon D t}$ where $C_i = \int_{\Omega_\epsilon} p_0(\boldsymbol{x}) w_{\lambda_i^\epsilon}(\boldsymbol{x}) d\boldsymbol{x} \int_{\Omega_\epsilon} w_{\lambda_i^\epsilon}(\boldsymbol{x}) e^{-\phi(\boldsymbol{x})} d\boldsymbol{x}$.

Starting with an equilibrium distribution $p_0(\boldsymbol{x}) = |\tilde{\Omega}|^{-1} e^{-\phi(\boldsymbol{x})}$, we have $C_i = |\tilde{\Omega}|^{-1}(\int_{\Omega_\epsilon} w_{\lambda_i^\epsilon}(\boldsymbol{x}) e^{-\phi(\boldsymbol{x})} d\boldsymbol{x})^2$ and finally the MFET is given by

$$\langle \tau_\epsilon \rangle = \sum_{i=0}^{\infty} \frac{C_i}{D\lambda_i^\epsilon}. \tag{94}$$

When the polymer distribution is sampled from the equilibrium distribution, we have that the significant contribution is coming from $C_0 \approx 1$, while the other terms are $C_i = o(1)$. Thus the first eigenvalue is the main contributor of the series.

4.5 Computing the eigenvalues of the Fokker-Planck equation and the MFET

The eigenvalues λ_i^ϵ of the operator $\mathcal{L}$ (eq.91) are obtained by solving the forward FPE in $\mathbb{R}^{d(N-1)}$, with the absorbing boundary condition on the boundary of the domain S_ϵ (see eq. 89), which is the tubular neighborhood of the $(N-1)d$-dimensional sub-manifold $\mathcal{S}$. For small ε, the eigenvalues can

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

be computed from a regular expansion near the eigenvalue of the entire space when the small domain S_ϵ is not removed:

$$\lambda_i^\epsilon = \lambda_i^0 + c_2\epsilon \int_S w_{\lambda_i^0}^2 dV_x + \mathcal{O}(\epsilon^2), \text{ for } d = 3 \tag{95}$$

$$\lambda_i^\epsilon = \lambda_i^0 + \frac{2\pi}{\log\epsilon} \int_S w_{\lambda_i^0}^2 dV_x + \mathcal{O}\left(\left(\frac{1}{\log\epsilon}\right)^2\right) \text{ for } d = 2, \tag{96}$$

where the eigenfunction $w_{\lambda_i^0}$ and eigenvalues λ_i^0 are associated to the non-perturbed operator (no boundary) [46], $d = 3, 2$. In the context of the Rouse polymer, the volume element is $dV_x = e^{-\phi(\boldsymbol{x})} d\boldsymbol{x}_g$, $d\boldsymbol{x}_g$, a measure over the sub-manifold $\mathcal{S}$ and $c_2 = \frac{2\pi^{3/2}}{\Gamma(3/2)}$ [46]. The unperturbed eigenfunctions $w_{\lambda_i^0}$ are products of Hermite polynomials [1], that depend on the spatial coordinates and the eigenvalues λ_i^0 are the sum of one dimensional eigenvalues [13].

The first eigenfunction associated to the zero eigenvalue is the normalized constant $w_{\lambda_0^0} = |\tilde{\Omega}|^{-1/2}$. The first eigenvalue for ε small is obtained from relation 95 in dimension 3 with $\lambda_i^0 = 0$ as the ratio

$$\lambda_0^\epsilon = \frac{c_2\epsilon \int_S e^{-\phi(\boldsymbol{x})} d\boldsymbol{x}_g}{|\tilde{\Omega}|} + \mathcal{O}(\epsilon^2), \tag{97}$$

of the closed loops to all polymer configurations. Using the expression for the potential ϕ (defined in 19), the volume is computed from Gaussian integrals

$$|\tilde{\Omega}| = \int_\Omega e^{-\phi(\boldsymbol{x})} d\boldsymbol{x}_g = \left[\frac{(2\pi)^{(N-1)}}{\prod_1^{N-1} \kappa_p}\right]^{d/2}, \tag{98}$$

while the parametrization of the constraint 90 leads to [13] after a direct computation

$$\int_S e^{-\phi(\boldsymbol{x})} d\boldsymbol{x}_g = \left[\frac{(2\pi)^{N-2} \prod_{p \text{ odd}} \omega_p^2}{\prod_p \kappa_p \left(\sum_{p \text{ odd}} \frac{\omega_p^2}{\kappa_p}\right)}\right]^{d/2}, \tag{99}$$

where $\omega_p = \cos(p\pi/2N)$. We now detail these computations: to evaluate the integration over S

$$|S|_{\lambda_0^\epsilon} = |\tilde{\Omega}|^{-1} \int_S e^{-\phi(\boldsymbol{x})} d\boldsymbol{x}_g, \tag{100}$$

note that S is embedded into $\mathbb{R}^{d(N-1)}$ by the immersion (for the detailed computations see [10, 14]),

$$\varphi : (\boldsymbol{t}_1, .., \boldsymbol{t}_{N-2}) \in \mathbb{R}^{d(N-2)} \to (\boldsymbol{t}_1, .., \boldsymbol{t}_{N-2}, -\sum_{p\text{odd}}^{N-2} \omega_p \boldsymbol{t}_p) \in \mathbb{R}^{d(N-1)}, \tag{101}$$

where

$$\omega_p = \frac{\cos(p\pi/2N)}{\sin(\pi/2N)}. \tag{102}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Thus $S = \varphi\left(\mathbb{R}^{d(N-2)}\right)$ is precisely the product of hyperplanes $\boldsymbol{t}_{N-1} = \varphi(\boldsymbol{t}) = -\sum_{p \text{ odd}}^{N-2} \omega_p \boldsymbol{t}_p$. The i-th element of the normal vector is

$$n^i = (\omega_1, 0, \omega_3, 0, ..., \omega_{N-1}) \tag{103}$$

and the normal vector to the submanifold is $\boldsymbol{n} = n^1 \otimes n^2 ... \otimes n^d$. Thus, the metric on S is induced by the restriction of the Euclidean distance

$$d\boldsymbol{x}_g \quad = \quad ||\boldsymbol{n}||d\boldsymbol{t} \tag{104}$$

and thus

$$|S|_{\lambda_0^\epsilon} = |\tilde{\Omega}|^{-1} \int_S ||\boldsymbol{n}|| \exp\left(-\frac{1}{2}\boldsymbol{t}\boldsymbol{B}\boldsymbol{t}^T\right) d\boldsymbol{t}, \tag{105}$$

where $\boldsymbol{B}$ is an $d(N-2) \times d(N-2)$ block matrix where the i-th block is given by

$$B^i_{pq} = \begin{cases} \kappa_p \delta_{pq} + \kappa_{N-1}\omega_p\omega_q, & p \text{ odd} \\ \kappa_p, & p \text{ even} \\ 0 & \text{otherwise.} \end{cases} \tag{106}$$

We use that this matrix can be re-written as

$$\boldsymbol{B}^i = \boldsymbol{C} + \kappa_{N-1}\boldsymbol{\omega} \otimes \boldsymbol{\omega}^T, \tag{107}$$

where $C_{pq} = \kappa_p \delta_{pq}$ and $\omega = (\omega_1, \omega_3, ..., \omega_{N-2})$. The determinant of $\boldsymbol{B}^i$ is computed using the matrix determinant theorem

$$\det(\mathrm{B_i}) = \prod_{\mathrm{p\ odd}} \kappa_{\mathrm{p}} \left(\sum_{\mathrm{p\ odd}} \frac{\omega_{\mathrm{p}}^2}{\kappa_{\mathrm{p}}}\right). \tag{108}$$

Finally, we get with

$$||\boldsymbol{n}||^2 = \sum_{p \text{ odd}} w_p^2, \tag{109}$$

that

$$|S|_{\lambda_0^\epsilon} = |\tilde{\Omega}|^{-1} \left[\frac{(2\pi)^{(N-2)} \sum_{p \text{ odd}} w_p^2}{\left(\sum_{p \text{ odd}} \frac{\omega_p^2}{\kappa_p}\right) \prod_p \kappa_p}\right]^{d/2}. \tag{110}$$

The numerator is

$$\sum_{i=0}^{m} \cos^2\left(\frac{(2i+1)\pi}{2N}\right) = \frac{m+1}{2} + \frac{\sin\frac{(2m+2)\pi}{N}}{4\sin\frac{\pi}{N}}; \ m = \begin{cases} \frac{N-2}{2}, & N \text{ even} \\ \frac{N-3}{2}, & N \text{ odd} \end{cases} \tag{111}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and the denominator

$$\sum_{p \text{ odd}}^{N-1} \cot^2\left(\frac{p\pi}{2N}\right) = \frac{N^2}{2} + \mathcal{O}(N). \tag{112}$$

Combining expressions 112 and 111 into equation 110, we get for large N

$$\lambda_0^\epsilon = \begin{cases} \left(\frac{\kappa}{N\pi}\right)^{3/2} 4\pi\epsilon + \mathcal{O}(\epsilon^2) & \text{for } d=3, \\ \frac{2\kappa}{N\log\left(\frac{b}{\epsilon}\right)} + \mathcal{O}\left(\left(\frac{1}{\log\epsilon}\right)^2\right) & \text{for } d=2. \end{cases} \tag{113}$$

This result shows in dimension three, for small ε, the MFET depends linearly on $\frac{1}{\varepsilon}$. In summary, for fixed N and small ε,

$$\lambda_0^\epsilon = \begin{cases} \left(\frac{\kappa}{N\pi}\right)^{3/2} 4\pi\epsilon + \mathcal{O}(\epsilon^2) & \text{for } d=3, \\ \frac{2\kappa}{N\log\left(\frac{\sqrt{2}b}{\varepsilon}\right)} + \mathcal{O}\left(\left(\frac{1}{\log\epsilon}\right)^2\right) & \text{for } d=2. \end{cases} \tag{114}$$

For small ε, the MFET depends linearly on $\frac{1}{\varepsilon}$ as confirmed by Brownian simulations (Fig. 14e). Because the zero order eigenvalue λ_0^ϵ is converging to the zero as ϵ tends to zero, it is well separated from the rest of the spectrum. Indeed, the second eigenvalues associated to the first two modes ($p = 1, 2$) are computed from the Hermite eigenfunctions [1], expressed with respect to the coordinates $(u_1^j, ..u_N^j)$ and $j = 1, ..d$

$$\begin{aligned} w_{\lambda^0_{1,j}}(u_1^j, ..u_N^j) &= \sqrt{\kappa_1}|\tilde{\Omega}|^{-1/2} u_1^j \text{ and} \\ w_{\lambda^0_{2,j}}(u_1^j, ..u_N^j) &= \sqrt{\kappa_2}|\tilde{\Omega}|^{-1/2} u_2^j \end{aligned}$$

where $\lambda_1^0 = \kappa_1$ and $\lambda_2^0 = \kappa_2$ respectively, which are the Rouse first mode (see relation 20). The expansion is [8]

$$\lambda_1^\epsilon \approx \kappa_1 + \epsilon c_2 \left(\frac{\kappa}{N\pi}\right)^{3/2} \left(1 - \frac{8}{\pi^2}\right) + \mathcal{O}(\epsilon^2), \tag{115}$$

$$\lambda_2^\epsilon \approx \kappa_2 + \left(\frac{\kappa}{N\pi}\right)^{3/2} 4\pi\epsilon + \mathcal{O}(\epsilon^2). \tag{116}$$

The contribution of λ_2^ϵ to the first looping time distribution is however quite small as shown in Fig. 14d (small deviation at the beginning), while λ_1^ϵ is not negligible. The MFET is well approximated by $\langle\tau_\epsilon\rangle \approx \frac{1}{D\lambda_0^\epsilon}$. When N increases, additional terms are needed in the expansion of λ_0^ϵ, while higher eigenvalues do not contribute significantly.

In summary, the zero eigenvalue is sufficient to characterize the MFET, confirming that the FET is almost Poissonian, except for very short time. Moreover, the second term in the expansion of λ_0^ϵ is proportional to $1/N$ [13]. Using the approximation $C_0 \sim 1$ and relation 94, for $d = 3$, the MFET is approximated by

$$\langle\tau_\varepsilon\rangle_{3d} \approx \frac{1}{D\lambda_0^\epsilon} = \frac{1}{D\left(\left(\frac{\kappa}{N\pi}\right)^{3/2} 4\pi\epsilon - A\epsilon^2/N\right)}, \tag{117}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where A is a constant that has been estimated numerically. Indeed, with $\epsilon = \frac{\varepsilon}{\sqrt{2}}$, the MFET is for d=3

$$\langle \tau_\varepsilon \rangle_{3d} = \left(\frac{N\pi}{\kappa}\right)^{3/2} \frac{\sqrt{2}}{D4\pi\varepsilon} + A_3 \frac{b^2}{D} N^2 + \mathcal{O}(1), \tag{118}$$

which hold for a large range of N, as evaluated with Brownian simulations (Fig. 17). The value of the coefficients are $A_3 = 0.053$ [13], obtained from fitting, comparable to the coefficient of N^2 obtained in [273], eq.13 ($A_3 = 0.053$, was estimated from the WF-approximation [287, 288]. These estimates are obtained for fixed N and small ε. Similar for $d = 2$, a a two dimensional space, the

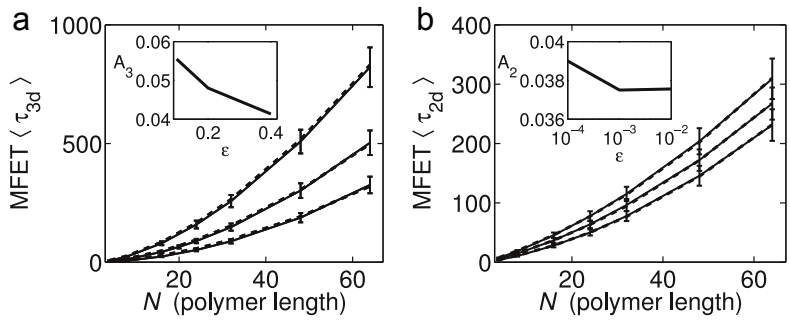


Fig. 17: **Mean first encounter time for different polymer lengths and different values of** ε. (**a**) MFET (three dimensions) estimated from Brownian simulations (full line) and compared to the theoretical MFET (eq.175) (dashed lines). The parameter A_3 is obtained by fitting ($\varepsilon = 0.1, 0.2, 0.4$). The unit is in b^2/D (as in fig. 14) (**b**) The MFET (two dimensions) estimated from Brownian simulation (full lines) and compared to the theoretical MFET (eq.174) (dashed lines). The parameter A_2 is obtained by fitting ($\varepsilon = 10^{-4}, 10^{-3}, 10^{-2}$) (reproduced from [13]).

asymptotic formula for MFET [13] is

$$\langle \tau_\varepsilon \rangle_{2d} = \frac{N}{2D\kappa} \log\left(\frac{\sqrt{2}b}{\varepsilon}\right) + A_2 \frac{b^2}{D} N^2 + \mathcal{O}(1), \tag{119}$$

All these asymptotic expansion are derived for fixed N and small ε. However, there should not be valid in the limit N large, although stochastic simulations (Fig. 17a-b) shows that the range of validity is broader than expected [13].

4.6 Looping in a confined domain

The looping time in confined geometries is computed following the same step as in free space, described in section 4.5. The terminology changes and we call this time the mean first encounter time in a confined domain (MFETC). The confined domain can be general but the computations consider the case of a ball. The principle of the computation is indeed to replace the reflecting boundaries for the monomers at the sphere by an external harmonic potential (Fig. 18).

We start with the Rouse polymer [72] containing N monomers, with a diffusion constant D, located inside a harmonic potential of strength B. Monomers are positioned at $\boldsymbol{R}_n$ ($n = 1, 2, ...N$). Their Brownian motion is coupled by a spring force to the nearest neighbors (see energy potential relation

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

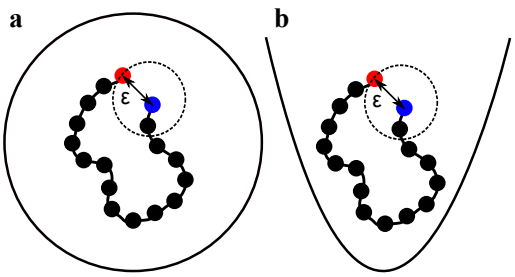


Fig. 18: **MFETC of polymer ends in a confined domain and in a potential well**. (**a-b**) Schematic representation of the encounter process between the ends of a polymer in a confined ball and in a harmonic potential.

5). The energy potential is the sum of the Rouse 5 plus a confined potential:

$$\phi_{\rm h}(\boldsymbol{R}_1,..\boldsymbol{R}_N) = \phi_{\rm Rouse}(\boldsymbol{R}_1,..\boldsymbol{R}_N) + \frac{B}{2}\sum_{n=1}^{N}\boldsymbol{R}_n^2 = \frac{1}{2}\sum_{p=0}^{N-1}(\kappa_p + B)\boldsymbol{u}_p^2, \tag{120}$$

where $\kappa_p = 4\kappa\sin^2(\frac{p\pi}{2N})$ and $\boldsymbol{u}_p$ are defined in 17. The dynamics of monomer $\boldsymbol{R}_n$ for $n = 1,..,N$ follows the stochastic equations

$$\frac{d\boldsymbol{R_n}}{dt} = -D\nabla_{\boldsymbol{R_n}}\phi_{\rm h} + \sqrt{2D}\frac{d\boldsymbol{w_n}}{dt}, \tag{121}$$

$\boldsymbol{w}_n$ are independent three-dimensional white Gaussian noises with mean zero and variance 1. The MFETC $\langle\tau_e\rangle$ is the mean time for the two ends $\boldsymbol{R}_N, \boldsymbol{R}_1$ to enter the distance ε, when the polymer evolves according to the new potential $\phi_{\rm h}$. The strength parameter B entering into the definition of the potential $\phi_{\rm h}$ is a free parameter, to be calibrated to match a confinement condition [8]: indeed, the parameter B is chosen such that the mean square end-to-end distance of the polymer is equal to the radius A of the confining ball, that is

$$\sqrt{\langle(\boldsymbol{R}_N(B) - \boldsymbol{R}_1(B))^2\rangle} = A \quad \text{for } N \gg 1, \tag{122}$$

By averaging over the Boltzmann distribution $e^{-\phi_{\rm h}}dx$, the explicit solution of eq. 122 is [8]

$$B = \frac{12}{A^4/b^2 + 2A^2}. \tag{123}$$

The asymptotic formula for the MFETC $\langle\tau_{\rm h}\rangle$ that two end monomers of a Rouse polymer moving in a harmonic potential meet is [10] for $\varepsilon \ll b$

$$\langle\tau_{\rm h}\rangle \approx \frac{2^{1/2}}{4\pi\varepsilon D}\left[\frac{4\pi N}{N^2B + \pi^2\kappa} + \frac{4}{\sqrt{\kappa\beta}}\left[\frac{\pi}{2} - \tan^{-1}\left(2\sqrt{\kappa/B}\tan\left(\pi/2N\right)\right)\right]\right]^{3/2} + \mathcal{O}(1). \tag{124}$$

This result is in contrast with previous formula derived for free looping polymers [202, 273, 82, 47, 271, 254, 13] (Fig. 19d): when N is large, the MFETC $\langle\tau_{\rm h}\rangle$ converges to a value $\frac{1}{\pi\varepsilon D}(\frac{\pi}{\sqrt{\kappa B}})^{3/2}$. The details asymptotic analysis is presented in the next section.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4.7 Detail computations of the looping time in a confined domain

Formula 182 [13] was derived for a Rouse polymer in confined domain, driven by eq.121 with the potential well $\phi_{\rm h}$ (see 120). The approximation starts with the relation $\langle \tau_{\rm h} \rangle \approx \frac{1}{D\lambda_0^\epsilon}$, where λ_0^ϵ is the first eigenvalue of the operator

$$\mathcal{L}p(\boldsymbol{u}) = \frac{D}{N}(\Delta_{u_0} p(\boldsymbol{u}) + \nabla_{u_0}(\nabla_0 \phi_{\rm h} p(\boldsymbol{u}))) + \sum_{k=1}^{N-1} D\Delta_{u_k} p(\boldsymbol{u}) + D\nabla_{u_k}(\nabla_p \phi_{\rm h} p(\boldsymbol{u})), \tag{125}$$

$\boldsymbol{u} = (\boldsymbol{u}_0, .., \boldsymbol{u}_{N-1}) \in \Omega = \mathbb{R}^{3N}$ and the absorbing boundary condition is $p(\boldsymbol{u}) = 0$ for $\boldsymbol{u} \in \partial S_\epsilon$, where S_ϵ is the ensemble of closed polymer configurations

$$S_\epsilon = \Big\{ \boldsymbol{u} = (\boldsymbol{u}_1, ..\boldsymbol{u}_{N-1}) \in \Omega \text{ s.t. } \sum_{p \text{ odd}} \boldsymbol{u}_p \cos(p\pi/2N) \leq \frac{\varepsilon}{2} \Big\}. \tag{126}$$

The first eigenvalue λ_0^ϵ is computed from the perturbation formula 95 (in dimension 3) [46]

$$\lambda_0^\epsilon = \lambda_0^0 + 4\pi\epsilon \int_{S_0} w_{\lambda_0^0}^2 dV_x, \tag{127}$$

where $w_{\lambda_0^0} = |\mathbb{R}^{3N}|_{\phi_{\rm h}}^{-1/2}$ is the constant eigenfunction associated with λ_0^0 and the volume element is $dV_x = e^{-\phi_{\rm h}} d\boldsymbol{x}_g$ and $d\boldsymbol{x}_g$ is the Euclidean measure over the sub-manifold S_0 (obtained by taking $\varepsilon = 0$ in eq.126). With $\lambda_0^0 = 0$ and a direct computation [9] leads to

$$\lambda_0^\epsilon = 4\pi\lambda_1^0 + \mathcal{O}(\epsilon^2), \tag{128}$$

where

$$\lambda_1^0 = \epsilon(2\pi)^{-3/2} \left[\frac{\sum_{p \text{ odd}} \omega_p^2}{K(N,B)} \right]^{3/2}, \tag{129}$$

$\omega_p = \cos(p\pi/2N)$ and

$$K(N,B) = \sum_{p \text{ odd}} \frac{\omega_p^2}{\kappa_p + B}. \tag{130}$$

For $N \gg 1$ relation $\langle \tau_\varepsilon \rangle \approx \frac{1}{D\lambda_0^\epsilon}$ and expanding relation 130 lead to equation 124 [9]. The second order in ϵ of λ_0^ϵ is estimated as follows [14, 48]:

$$\lambda_0^\epsilon \sim 4\pi\lambda_1^0\epsilon + (4\pi)^2\lambda_2^0\epsilon^2 + \mathcal{O}(\epsilon^3), \tag{131}$$

with $\lambda_2^0 = \tilde{\gamma} \int_{S_0} \int_{S_0} G(\boldsymbol{x};\boldsymbol{y}) dV_x dV_y$, where $\tilde{\gamma}$ is a constant that depends on the geometry of the boundary. This constant measures the contribution of the second order term in the expansion of the MFET. The Green's function G is associated to the operator $\mathcal{L}$, solution of $\mathcal{L}\mathcal{G} = -\delta$, where δ is the Dirac operator. The Green's function is expanded on the eigenfunctions $w_{\lambda_i^0}$ and eigenvalues λ_i of the operator $\mathcal{L}$, so that

$$G(\boldsymbol{x};\boldsymbol{y}) = -\sum_{i \neq 0} \frac{w_{\lambda_i^0}(\boldsymbol{x}) w_{\lambda_i^0}(\boldsymbol{y})}{\lambda_i}, \tag{132}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

which leads to the approximation [14]

$$\lambda_2^0 = -\tilde{\gamma} \sum_{k \neq 0} \frac{\langle w_{\lambda_k^0} | w_{\lambda_0^0} \rangle_{S_0}^2}{\lambda_k} \approx \frac{\langle w_{\lambda_2^0} | w_{\lambda_0^0} \rangle_{S_0}^2}{\lambda_2}, \tag{133}$$

where

$$\langle w_{\lambda_i^0} | w_{\lambda_0^0} \rangle_S = \int_{S_0} e^{-\phi_{\mathrm{h}}(\boldsymbol{u})} w_{\lambda_i^0}(\boldsymbol{u}) w_{\lambda_0^0}(\boldsymbol{u}) dV_g. \tag{134}$$

Since the first non-zero eigenfunction ($w_{\lambda_{1,i}^0}$) is linear with the coordinates $u_{1,i}$, the scalar product is zero and equation 133 is computed from the product associated with the second eigenvalue and $p = 1$ in the spatial directions i. The eigenfunctions are given by the second Hermite polynomial

$$w_{\lambda_{2,i}^0}(u_{1,i}) = (2|\mathbb{R}^{3N}|_{\phi_{\mathrm{h}}})^{-1/2} \left((\kappa_1 + B)(u_{1,i})^2 - 1 \right). \tag{135}$$

and $\lambda_2 = 2(\kappa_1 + B)$. A direct computation gives [14]

$$\lambda_0^\epsilon = \frac{\varepsilon}{8\pi^{1/2}} \left[\frac{N}{K(N,B)} \right]^{3/2} - \frac{2^{-8}\pi^{-1}\varepsilon^2 \tilde{\gamma} N^3}{(\kappa_1 + B)^3 K(N,B)^5}, \tag{136}$$

where

$$K(N,B) = \sum_{p \text{ odd}} \frac{\omega_p^2}{\kappa_p + B}. \tag{137}$$

For large N, $K(N,B) \approx \frac{N}{4\sqrt{\kappa B}}$ and $\kappa_1 = 4\kappa \sin^2(\frac{\pi}{2N})$. Finally, using $\langle \tau_\varepsilon \rangle \approx \frac{1}{D\lambda_0^\epsilon}$ for small ε and N fixed, the approximation 127 is refined, leading to the MFETC expression

$$\langle \tau_\epsilon \rangle \approx \frac{2^3 \pi^{1/2}}{D\varepsilon} \left[\frac{K(N,B)}{N} \right]^{3/2} + \frac{\tilde{\gamma}(16\kappa \sin^2(\frac{\pi}{2N}) + B)^{-3}}{DK(N,B)^2}.$$

Contrary to the case of a free space (eq. 118) where $\langle \tau_\epsilon \rangle = a_1 N^{3/2}/\varepsilon + a_2 N^2$ (a_1, a_2 are constants) and the N^2 term dominates for $N \gg 1$, in the confined case, only the first term is increasing with the length N and the two asymptotic limit is bounded by the diameter of the domain.

The looping time distribution in confined domains is Poissonian. It is also the survival probability $P(t)$ that a loop is not formed before time t (see section 4.4). In this approximation,

$$P(t) = \sum_{i=0}^{\infty} C_i e^{-\lambda_i^\epsilon D t} \approx e^{-\lambda_0^\epsilon D t}, \tag{138}$$

where C_i are coefficients. Brownian simulations of the two ends of a polymer in a harmonic well (Fig. 19a) confirms this Poissonian approximation. Unlike the FET distribution for a free polymer, $P(t)$ is well approximated by a single exponential, even for long polymers, showing that the higher exponential terms (eq.138) do not contribute.

In summary, the encounter time for a polymer in a harmonic well is Poissonian. Moreover, as observed from matching the theoretical formula 182 and Brownian simulations for the MFETC in a potential well (Fig. 19b), the first order correction in ε is enough and this approximation remains valid for large N and small ε. The true formula for a bounded domain, without using the parabolic confinement is still to be found.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

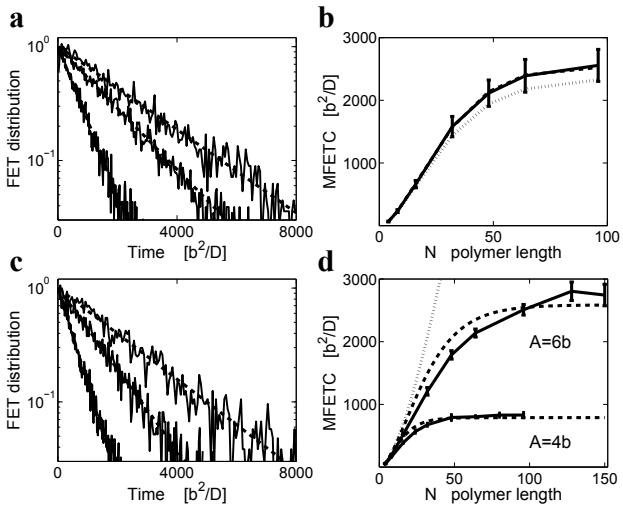


Fig. 19: **Encounter of the two monomer ends for confined Rouse polymer**. (**a**) First encounter time distributions for a Rouse polymer (eq. 14) confined in a harmonic well for various polymer lengths $N = 16, 32, 64$ (left to right) with $B = 0.01b^{-2}$ and $\varepsilon = 0.01b$. A single exponential (dashed line) is enough to describe the process. (**b**) Brownian simulations (full line) are compared to the MFETC computed eq.128 by taking the first order only in ε (pointed line) and by taking into account the second order correction (dashed line) given by eq.136. The value $\tilde{\gamma} = 10$ is obtained by fitting expression 138 to stochastic simulations. (**c**) First encounter time distributions for a polymer in a ball of radius $A = 6b$, for various polymer lengths ($N = 16, 32, 64$ left to right). A single exponential (dashed line) is sufficient to approximate the distribution over the entire range. (**d**) MFETC as a function of the polymer length: Brownian simulations (full lines) in spheres of radii $A = 4b, 6b$ with $\varepsilon = 0.01b$. The MFETC (dashed lines) is estimated by using an harmonic well approximation eq.128, where B was fitted to the simulation results $B_4 = 0.0406b^{-2}, B_6 = 0.0089b^{-2}$. The MFET is also shown (points) for a freely moving polymer (eq.2 [13]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4.8 Brownian simulations of looping in a confined domain

The range of validity of formula 182 for a polymer looping in a confined domain can be estimated using Brownian simulations of a Rouse polymer in a ball (Fig. 19d) for different radii. Although for small polymers, there is almost no differences between confined and non-confined looping time, for long polymers, the MFETC reaches an asymptotic value that depends on the radius of the domain. When the interacting monomers are not at the ends of polymer chain, the MFETC is reduced in a range between five to thirty percent, due to the interactions of the additional monomers with the boundary [8].

The distribution of arrival times (Fig. 19c) for short and long polymers is well approximated by a single exponential, confirming that the looping event is almost Poissonian. This is in contrast with looping in free space, where for longer polymers, a second exponential is necessary. The MFETC is computed by taking into account only the first order terms in ε (see eq.128). The strength B is estimated by comparing the analytical formula with Brownian simulations: the calibration formula 123 gives similar values to the fitted ones. For $A = 4b, 6b$, the fitted values are $B_{\text{fit},4} = 0.0406b^{-2}, B_{\text{fit},6} = 0.0089b^{-2}$ respectively, while the calibration formula 123 gives $B_{\text{cal},4} = 0.0417b^{-2}, B_{\text{cal},6} = 0.0088b^{-2}$. Using the fitting procedure (formula 123), it is possible to estimate the radius of a confined sphere. The MFETC has different applications such as estimating the chromosomal interaction forces, the nuclear sub-organization [157, 110] and the encounter between two sites that can initiate gene activation or regulation [66, 209, 76].

In summary, the FLT is well approximated by a single exponential, showing that the associated stochastic process is almost Poissonian. Consequently, looping in the nucleus, chromosome or telomere encounter can be well characterized by the mean looping time. This approximation simplifies heavy numerical simulations or reduce analysis to Poissonian processes, such the analysis of telomere clustering in the yeast nucleus [122]. Another conclusion is that by increasing the radius ε or the polymer length N, the asymptotic formula for the MFET is completely determined by the first eigenvalue, but not by higher order eigenvalues. Two scales are involved in the MFET, one proportional to N^2 and the other to $N^{3/2}$ and both are already contained in the first eigenvalue and do not arise from higher ones. It is surprising that the regular perturbation of the Fokker-Planck operator in ε (formula 95) introduces a novel scale with N in all eigenvalues. A complete expansion for the MFET has to be found and it would be interesting to find a geometrical characterization of the constants A_2 and A_3.

4.9 Looping formula for rod-like long polymer

In this subsection, we analyze a different polymer model called the rod-like model (introduced in paragraph 3.9) which is more elementary than Rouse, but this model is not really used for describing looping. The reason is that this model shows correlation of the last monomer with the rest of polymer chain (see [12] for a complete discussion). However, the analysis of this model is quite simple and contains interesting asymptotic features.

We shall now describe the MFET computation for the rod-like polymer. The analysis starts with approximating the stochastic process of the position $\boldsymbol{R}_N(t)$ in the large N limit in dimension 2 [12]. Ito's formula for the exponential of the Brownian angle θ_k gives

$$de^{i\theta_k(t)} = ie^{i\theta_k(t)}d\theta_k(t) - De^{i\theta_k(t)}dt. \tag{139}$$

By summing relations 139, using relation 41 for $\boldsymbol{R}_k(t) = b\sum_{j=1}^{k}\exp(i\theta_j(t))$, we get

$$d\boldsymbol{R}_N(t) = -D\boldsymbol{R}_N(t)dt + bd\tilde{\boldsymbol{W}}_N(t), \tag{140}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where the new noise term is $d\tilde{\boldsymbol{W}}_N(t) = i\sum_{k=1}^N e^{i\theta_k(t)}d\theta_k(t)$. The noise source $\tilde{\boldsymbol{W}}_N(t)$ can be approximated by a Brownian motion for large N in the following sense: $\frac{\tilde{\boldsymbol{W}}_N(t)}{\sqrt{N}}$ converges asymptotically (in probability) to a Brownian motion [278]. While the first moment of $\tilde{\boldsymbol{W}}_N$ is zero, the second order is computed from the decomposition $d\tilde{\boldsymbol{W}}_N(s) = (d\tilde{W}_N^1(s), d\tilde{W}_N^2(s))$, where

$$d\tilde{W}_N^1(s) = -\sum_{k=1}^N \sin\theta_k(t)d\theta_k(t), d\tilde{W}_N^2(s) = \sum_{k=1}^N \cos\theta_k(t)d\theta_k(t), \tag{141}$$

are derived from $\langle d\theta_j(s)d\theta_k(s)\rangle = 2D\delta_{j,k}ds$. Indeed,

$$\begin{aligned}\int_0^t \langle d\tilde{W}_N^1(s), d\tilde{W}_N^1(s)\rangle &= \int_0^t \sum_{j,k=1}^N \sin\theta_j(s)\sin\theta_k(s)\langle d\theta_j(s)d\theta_k(s)\rangle = 2DE\left(\int_0^t \sum_{k=1}^N \sin^2(\theta_k(s))\right)ds \\ &= 2NDE\left(\int_0^t \sin^2(\theta_1(s))\right)ds = 2ND\left(\frac{t}{2} + \frac{1-e^{-4Dt}}{8D}\right),\end{aligned}$$

because $E\left(\sin^2(\theta_1(t))\right)ds = \frac{1-e^{-4Dt}}{2}$ and $\int_0^t E\left(\sin^2(\theta_1(t))\right)ds = \frac{t}{2} + \frac{1-e^{-4Dt}}{8D}$. Similarly,

$$\begin{aligned}\int_0^t \langle d\tilde{W}_N^2(s), d\tilde{W}_N^2(s)\rangle &= \int_0^t \sum_{j,k=1}^N \cos\theta_j(s)\cos\theta_k(s)\langle d\theta_j(s)d\theta_k(s)\rangle = 2ND\left(\frac{t}{2} + \frac{1+e^{-4Dt}}{8D}\right) \approx NDt, \\ \int_0^t \langle d\tilde{W}_N^1(s), d\tilde{W}_N^2(s)\rangle &= \int_0^t \sum_{j,k=1}^N \sin\theta_j(s)\cos\theta_k(s)\langle d\theta_j(s)d\theta_k(s)\rangle = 0\end{aligned}$$

For large times, the leading order term for the second order moments is NDt. Thus to compute the MFET of the end-points of the long polymer, the noise term $d\tilde{W}_N^i$ was approximated by a Brownian motion in both x- and y-component, with a diffusion constant ND and the stochastic equation 140 is approximated by the following Ornstein-Ulhenbeck process

$$d\boldsymbol{R}_N(t) = -D\boldsymbol{R}_N(t)dt + \sqrt{2DN}bd\boldsymbol{W}(t), \tag{142}$$

where $\boldsymbol{W}$ is a two-dimensional Brownian motion of variance one. The MFET can be computed by considering the dynamics inside a ring with exterior and interior radii Nb and ϵ respectively. The associated partial differential equation is described in 144, where the inner boundary $r = \epsilon$ is absorbing, while a Neumann boundary condition is imposed on the external boundary, the looping time $u(r)$, where initially $|\boldsymbol{R}_N| = r$ is solution of the partial differential equations

$$\begin{aligned}NDb^2\Delta u - D < X, \nabla u > &= -1 \text{ on } D(0, Nb), \\ u(P) &= 0 \text{ for } |P| = \epsilon b, \\ \frac{\partial u}{\partial n}(P) &= 0 \text{ for } |P| = Nb,\end{aligned} \tag{143}$$

where X is the radial vector, $D(0, Nb)$ is the disk centered at the origin of radius $R = Nb$. The equation is solved in polar coordinates $NDb^2\left(u'' + \frac{u'}{r}\right) - Dru' = -1$ on $D(0, Nb)$, $u(\epsilon) = 0, u'(Nb) = 0$, and the solution is

$$u(r) = \frac{1}{D}\log\left(\frac{r}{\epsilon b}\right) - \frac{1}{D}e^{\frac{-N}{2}}\int_{\epsilon b}^r \frac{e^{\frac{v^2}{2Nb^2}}}{v}dv. \tag{144}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Interestingly, when the initial configuration of the rod-polymer is a straight segment ($r = R$), the MFLT is $u(R) = \frac{1}{D}\log\left(\frac{R}{\epsilon}\right) - \frac{1}{D}e^{\frac{-N}{2}}\int_{\epsilon b}^{R}\frac{e^{\frac{v^2}{2Nb^2}}}{v}dv$. Integrating by parts and using Laplace's method for N large, the MFET is $u(R = Nb) \approx \frac{1}{D}\log\left(\frac{1}{\epsilon}\right)\left(1 - e^{-\frac{N}{2}+\frac{\epsilon^2}{2N}}\right) + \frac{1}{D}\log(N)\left(1 - e^{(\epsilon - N)}\right)$. Finally,

$$\langle \tau_\epsilon \rangle = \frac{1}{D}\log\left(\frac{N}{\epsilon}\right) + \mathcal{O}_{\epsilon,\mathcal{N}}(1). \tag{145}$$

Three regimes can be distinguished depending whether N is $\gg \ll$ or $\approx \frac{1}{\epsilon}$. When $N \gg \frac{1}{\epsilon}$, the dominant behavior depends on $\log(N)$. The MFLT for a chain starting in a uniform distribution is

$$\langle \tau_\epsilon \rangle = \frac{1}{\pi R^2}\int_{\epsilon b}^{R} u(r) r dr \approx \frac{1}{2\pi D}\log\left(\frac{N}{\epsilon}\right) + \mathcal{O}_{\epsilon,\mathcal{N}}(1). \tag{146}$$

This asymptotic formula is different from the one obtained for a chain initially straight (formula 145). The leading order term depends on the initial configuration of the polymer. This is quite surprising because the prediction of the narrow escape formula for a Brownian particle implies that the leading order term does not depend on the initial configuration [239]. In summary, the initial configuration for a rod-like polymer affects the leading order asymptotic expansion of the MFLT, suggesting that the underlying stochastic process is not Markovian and has a long memory, a phenomena that should be further investigated.

When the initial condition is chosen at equilibrium with a distribution $p(r) = Pr\{R_N = r\} = \frac{1}{2\pi N b^2}e^{-\frac{r^2}{2Nb^2}}$, then the MFLT is given by

$$\langle \tau_\epsilon \rangle = \frac{1}{D}\log\left(\frac{\sqrt{2N}}{\epsilon}\right) + \mathcal{O}\left(\frac{1}{N}\right) + \mathcal{O}(1). \tag{147}$$

The leading order term of the MFLT depends again on the initial configuration and should be compared with formula 146.

4.10 Looping time for a $4^{th}-$order stiff polymer using a non-Markovian approach

The dynamics of a single monomer from a polymer chain and the mean first looping time was considered using Brownian simulations using a non-Markovian approach [99, 100]. In that approach, the parameter of the model are the length (the number of monomer N), the same diffusion coefficient D for each monomer, the bond length l_0, a target size a such that $a \ll \frac{l_0}{\sqrt{N}}$ inside a domain of volume V. In that case, the leading order term for the mean time recover the classical result of a search to a small target

$$\langle \tau \rangle \approx \sqrt{\pi/8}\, l_0^3 N^{3/2}\frac{1}{Da}, \tag{148}$$

while for $l_0 N \gg a \gg \frac{l_0}{\sqrt{N}}$,

$$\langle \tau \rangle = \frac{VN}{4\pi D a_e}, \tag{149}$$

where the effective radius is defined by $a_e = l_0\sqrt{N}$, and does not depend on the target size (when a is very small). The general formula is derived in section 4.6.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Interestingly, the mean looping time $\langle\tau\rangle$ for a semi-flexible Gaussian chain can be estimated in a polymer model, where the local interactions tends to align successive bonds. A continuous model is discussed in [107, 90] for a stiff polymer. The tension along the chain is approximated by a constant g, neglecting the long-range hydrodynamic, b_0 is the thickness of the chain, κ is the bending rigidity. The equation of motion for a segment $r(s,t)$ located at position s at time t is

$$\eta\frac{\partial r(s,t)}{\partial t} = -b_0\kappa\frac{\partial^4 r(s,t)}{\partial s^4} + 2g\frac{\partial^2 r(s,t)}{\partial s^2} + \sqrt{2\eta k_B T}\frac{dw}{dt} \tag{150}$$

where η is the friction per unit length and w the classical normalized Gaussian noise in time and delta-correlated in space. Four reflective boundary conditions are imposed at the boundary of the chain [101]. The evolution of the position can be computed in the Fourier space from an ensemble of OU-processes as follow

$$\eta_p\frac{dX_p(t)}{dt} = -k_p X_p(t) + \sqrt{2\eta k_B T}\frac{dw_p}{dt} \tag{151}$$

for $p \neq 0$, $k_p = ap^4 + bp^2$, $a = 2\kappa b_0(\pi/L)^4, b = 4g(\pi/L)^2$, w_p are independent normalized Brownian motion, with

$$X_p(t) = \frac{1}{L}\int_0^L r(s,t)\cos(p\pi s/L)ds. \tag{152}$$

The looping time of a semi-flexible chain when the smallest length scale is the bending fluctuation length and the target size is $a \ll 1$ has been estimated asymptotically

$$\langle T\rangle = Ca^{-1/3}, \tag{153}$$

where C is constant [101]. This scaling law is a characterization of looping for a stiff polymer. The regime where $\kappa \gg l_0 N$ should be very different from the one discussed in this section and the asymptotic analysis should be clarified.

4.11 Mean First Encounter Time (MFET) between two monomers located on the same $\beta-$ polymer

The MFET between two monomers of a $\beta-$polymer (see section 3.5) is computed from the first eigenvalue expansion of the associated Fokker-Planck operator. We proceed now following the same steps as in the previous paragraph, neglecting the contributions of all higher eigenvalues, thus

$$\langle\tau_\epsilon^\beta\rangle \approx \frac{1}{D\lambda_0^\epsilon}. \tag{154}$$

To estimate λ_0^ϵ to first order in ε, we use formula [14]

$$\lambda_0^\epsilon = \frac{4\pi\epsilon\int_{\text{C-P}} e^{-\phi(\boldsymbol{x})}d\boldsymbol{x}_g}{|\tilde{\Omega}|} + \mathcal{O}(\epsilon^2), \tag{155}$$

which is the ratio of a Gaussian integrals over all closed polymer (C-P) to the whole polymer configuration ensemble. Integral 155 is computed directly. Indeed, using the expression for the potential 25, we get

$$|\tilde{\Omega}| = \int e^{-\phi(\boldsymbol{x})}dV_{\boldsymbol{x}} = (2\pi)^{(N-1)d/2}\left(\prod_{p=1}^{N-1}\tilde{\kappa}_p\right)^{-3/2}. \tag{156}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The numerator of equation 155 involves integrating over the ensemble of closed polymer loops

$$|S|_{\lambda_0^\epsilon} = |\tilde{\Omega}|^{-1} \int_{C-P} e^{-\tilde{\phi}(\boldsymbol{x})} d\boldsymbol{x}_g, \tag{157}$$

where $d\boldsymbol{x}_g$ is the induced metric of the configuration space onto the set C-P. A direct integration of eq. 157 gives (see paragraph 4.5) [14]

$$|S|_{\lambda_0^\epsilon} = |\tilde{\Omega}|^{-1} \left[\frac{(2\pi)^{(N-2)} \sum_{p \text{ odd}} w_p^2}{\left(\sum_{p \text{ odd}} \frac{\omega_p^2}{\tilde{\kappa}_p} \right) \prod_p \tilde{\kappa}_p} \right]^{d/2} = |\tilde{\Omega}|^{-1} \left[\frac{(2\pi)^{(N-2)} S_1(N)}{\frac{1}{4\kappa} S_2(N) \prod_p \tilde{\kappa}_p} \right]^{d/2}, \tag{158}$$

where $\omega_p = \cos(p\pi/2N)$ and the series

$$S_1(N) = \sum_{i=0}^{m} \cos^2\left(\frac{(2i+1)\pi}{2N}\right) = \frac{m+1}{2} + \frac{\sin\frac{(2m+2)\pi}{N}}{4\sin\frac{\pi}{N}}; \; m = \begin{cases} \frac{N-2}{2}, & N \text{ even} \\ \frac{N-3}{2}, & N \text{ odd} \end{cases} \tag{159}$$

and

$$S_2(N) = \sum_{p \text{ odd}}^{N-1} \frac{\cos^2\left(\frac{p\pi}{2N}\right)}{\sin^\beta\left(\frac{p\pi}{2N}\right)} \tag{160}$$

is approximated by the Euler-Maclaurin formula by

$$S_2(N) \approx \int_0^{\frac{N-1}{2}} g(q,N,\beta)dq + \frac{g(\frac{N-1}{2},N,\beta) + g(0,N,\beta)}{2} + \frac{1}{12}(g'(\frac{N-1}{2},N,\beta) - g'(0,N,\beta)), \tag{161}$$

where $g(p,N,\beta) = \frac{\cos^2\left(\frac{(2p+1)\pi}{2N}\right)}{\sin^\beta\left(\frac{(2p+1)\pi}{2N}\right)}$ and the integral is

$$\int_0^{(N-1)/2} g(s,N,\beta)ds = \frac{N}{3\pi}\cos\left(\frac{\pi}{2N}\right)^3 {}_2F_1\left(\frac{3}{2}, \frac{1+\beta}{2}, \frac{5}{2}, \cos\left(\frac{\pi}{2N}\right)^2\right), \tag{162}$$

where ${}_2F_1$ is the Gaussian hypergeometric function [1]. The terms $g(0,N,\beta) = \frac{\cos^2\left(\frac{\pi}{2N}\right)}{\sin^\beta\left(\frac{\pi}{2N}\right)}$ and $g'(0,N,\beta) = -\frac{\beta\pi\cos\left(\frac{\pi}{2N}\right)^3}{2N\sin(\frac{\pi}{2N})^{1+\beta}}$, in 161 are of order $\mathcal{O}(N^\beta)$. The series $S_2(N)$ is approximated by

$$S_2(N) \approx \frac{N}{3\pi}\cos\left(\frac{\pi}{2N}\right)^3 {}_2F_1\left(\frac{3}{2}, \frac{1+\beta}{2}, \frac{5}{2}, \cos\left(\frac{\pi}{2N}\right)^2\right) + \frac{1}{2}\frac{\cos^2\left(\frac{\pi}{2N}\right)}{\sin^\beta\left(\frac{\pi}{2N}\right)} + \frac{\pi}{24N}\frac{\beta\cos\left(\frac{\pi}{2N}\right)^3}{\sin(\frac{\pi}{2N})^{1+\beta}}. \tag{163}$$

Thus substituting relations 163,159, 158 and 156 into eq. 155, we get for large N,

$$\lambda_0^\epsilon \approx 4\epsilon\kappa^{-3/2}\left[\frac{2}{3}\cos\left(\frac{\pi}{2N}\right)^3 {}_2F_1\left(\frac{3}{2}, \frac{1+\beta}{2}, \frac{5}{2}, \cos\left(\frac{\pi}{2N}\right)^2\right) + \frac{\pi}{N}\frac{\cos^2\left(\frac{\pi}{2N}\right)}{\sin^\beta\left(\frac{\pi}{2N}\right)} + \frac{\pi^2}{12N^2}\frac{\beta\pi\cos\left(\frac{\pi}{2N}\right)^3}{2\sin(\frac{\pi}{2N})^{1+\beta}}\right]^{-3/2}. \tag{164}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For large N, ${}_2F_1\left(\frac{3}{2}, \frac{1+\beta}{2}, \frac{5}{2}, \cos\left(\frac{\pi}{2N}\right)^2\right) = 3\pi^{3/2-\beta}2^{-(1+\beta)}N^{\beta-1} + \mathcal{O}(1) + \mathcal{O}(N^{-2})$, while the two other terms in the parenthesis scale like $N^{\beta-1}$. Finally, the mean encounter time for the two end monomers is

$$\langle \tau_\epsilon^\beta \rangle \approx \frac{1}{D\varepsilon(2\kappa)^{3/2}} \left[\frac{2}{3}\cos\left(\frac{\pi}{2N}\right)^3 {}_2F_1\left(\frac{3}{2}, \frac{1+\beta}{2}, \frac{5}{2}, \cos\left(\frac{\pi}{2N}\right)^2\right) + \frac{\pi}{N}\frac{\cos^2\left(\frac{\pi}{2N}\right)}{\sin^\beta\left(\frac{\pi}{2N}\right)} + \frac{\pi^2}{12N^2}\frac{\beta\pi\cos\left(\frac{\pi}{2N}\right)^3}{2\sin(\frac{\pi}{2N})^{1+\beta}} \right]^{3/2}.$$

In summary, the MFETs are quite different for a Rouse or a β-polymer: for the latter, it scales with $N^{3/2}$, while for a β-polymer in the limit $N \gg 1$, it behaves like $N^{\frac{3}{2}(\beta-1)}$.

4.12 Search for a small target located on a boundary

In previous sections, we studied the looping time and we now focus on the search process by single monomer for a small target located on the boundary of a ball (nucleus). In a biological context, targets can be small nuclear pores located on the nuclear envelope or active genes that need to be found and activated [5]. Finding a small nuclear pore can also arise in the context of gene delivery, where RNA fragments have to enter the nucleus pores. During the repair of a double-stranded DNA repair, a DNA end has to search for its other broken end in the confined microdomain generated by the chromatin environment [165] and can also re-localize to the membrane periphery to interact with a nuclear pore.

As there are no analytical formula available for this search process, we present several coarse-grained numerical results of the first arrival time of a monomer from polymer to a small target, called the narrow encounter time for a polymer NETP. In that case, the small target is located on the surface of a bounded microdomain. The polymer model we consider here is a Freely-Joint-Chain [273] (monomers connected by springs with a none-zero resting length l_0).

There are several possibilities to state the search problem:

1. any one of the monomer or
2. only one

can find the small target. Simulations reveal that the NETP is an increasing function of the polymer length until a critical length is reached, passed this length, it decreases (see Fig. 20). Interestingly, for the second case, the position of the searching monomer along the polymer that can be absorbed strongly influences the NETP. Computing the NETP is relevant because it is the reciprocal of the forward activation rate for diffusion limited processes of chemical reactions for a site located on a polymer [299, 202]. It may also be used to estimate the first time that a gene is activated by a factor located on the DNA, which is different from the classical activation due to a transcription factor [251, 168, 24, 215]. However, at this stage a final satisfactory analytical formula is still missing (see [117] for an attempt).

We now recall some results of Brownian simulations for the two cases mentioned above:

1. When *any* of the polymer monomer can be absorbed at the target region, the search time is designated by $\langle \tau_{\text{any}} \rangle$.
2. When only one fixed monomer can find the target region, it is $\langle \tau_{\text{mon}} \rangle$.

The key parameters that influence the search time are the polymer length, the monomer's position along the chain and the bending elasticity [8]. In dimensions two and three, the NETP increasing with

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the polymer length (number of monomers N), until a critical value N_c after which it is decreasing. The radius of gyration $R_g = \sqrt{N_c} l_0/6$ (where l_0 is defined as the shortest length between the nearest monomers at equilibrium $< |\boldsymbol{x}_k - \boldsymbol{x}_{k+1}| >= l_0$) can be used to characterize the coupling between the polymer and the boundary of the microdomain. For short polymer lengths, such that $R_g \ll 2R$

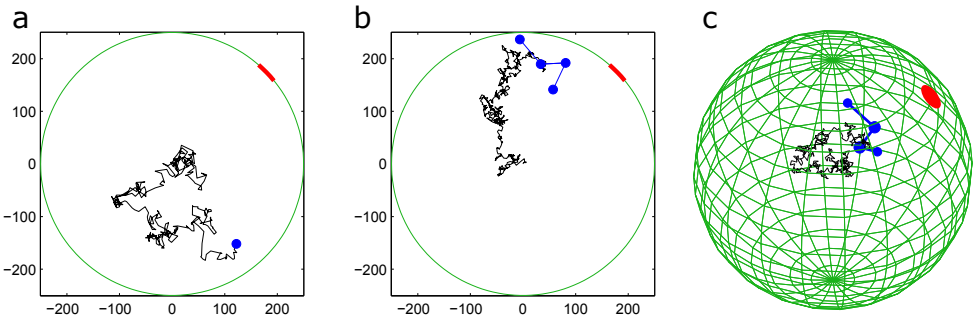


Fig. 20: **Confined trajectories of FJC polymer.** In a circular(2D)/spherical(3D) domain of radius 250 nm, the target size is $\varepsilon = 50$nm. Trajectories of the center of mass (black) before absorption at the target (red). **(a)** Simulation with one monomer, moving in two dimensions. **(b-c)** Simulations of a 4-bead Rouse polymer, moving in 2 dimensions (b) and in 3 dimensions (c).

(diameter of the microdomain), the NETP is largely determined by the motion of the center of mass. When the polymer is far from the absorbing target, none of the monomers will be able to reach it, until the center of mass has moved close to the target. The NETP thus reflects the mean first passage time of the center of mass to the target. In this limiting case, the center of mass undergoes Brownian motion with a diffusion constant inversely proportional to the number of monomers: $D_{\rm CM} = \frac{D}{N}$. Thus, in the regime $N \ll N_c$, the NETP is approximately that of a single Brownian particle, but with a smaller diffusion constant. The expressions for the NET are inversely proportional to $D_{\rm CM}$, and thus we obtain the initial linear regime in N:

$$\begin{aligned} \langle \tau_{\rm any} \rangle_{2d} &\approx N \langle \tau_{2d} \rangle \\ \langle \tau_{\rm any} \rangle_{3d} &\approx N \langle \tau_{3d} \rangle, \end{aligned} \tag{165}$$

as confirmed by numerical simulations in Fig. 21. However for a polymer of length comparable to the size of the microdomain, the location of the center of mass does not determine anymore the NETP. In that regime, smaller subsections of the polymer can be close to the target even if the center of mass is far away. In addition, when any bead can be absorbed, increasing the polymer length results in a decrease in the NETP (Fig. 21). Interestingly, two regimes can be further distinguished for the decay phase. When any bead can be absorbed, the decay phase of the NEPT (Fig 21) can be separated into two different regimes that can be described as followed. In the first one, the polymer moves freely until a monomer hits the absorbing boundary. The NETP is determined by the competitive effects of a decreased diffusion constant for the center of mass and an increased total polymer length. Increasing the polymer length leads to an effective smaller volume of the effective confining domain in which the polymer has to find the absorbing window.

When the length of the polymer becomes long enough, so that at least one monomer can always be found in the boundary layer (of size ε) of the absorbing hole [239], the NETP has a different decay as a function of N compared to the previous intermediate regime. In that case, the center of mass is strongly restricted due to the interaction of all the monomers with the microdomain surface (Fig. 22). The NETP is determined by mean time for a monomer in the boundary layer of the absorbing

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

window. It is still an unsolved problem to obtain an asymptotic estimate for that time. A fitting procedure shows that the NETP scale with two exponentials

$$\frac{\langle \tau_{\rm any}(N)\rangle_{2d}}{\langle \tau_{2d}\rangle} = a_2 \exp(-\alpha_2 N) + b_2 \exp(-\beta_2 N) \text{ in dimension } 2 \tag{166}$$

$$\frac{\langle \tau_{\rm any}(N)\rangle_{3d}}{\langle \tau_{3d}\rangle} = a_3 \exp(-\alpha_3 N) + b_3 \exp(-\beta_3 N) \text{ in dimension } 3. \tag{167}$$

$\langle \tau_{3d}\rangle$ and $\langle \tau_{2d}\rangle$ are the NET for a single particle τ_0, in dimension 3 and 2 respectively. In dimension two, the exponents are $\alpha_2 = 0.0075, \beta_2 = 0.024$ and coefficients $a_2 = 0.23, b_2 = 2.17$ and in dimension 3, $\alpha_3 = 0.0082, \beta_3 = 0.030$ and coefficient $a_3 = 0.60, b_3 = 1.56$ (Fig. 21). The empirical laws 166-167 have not yet been derived analytically.

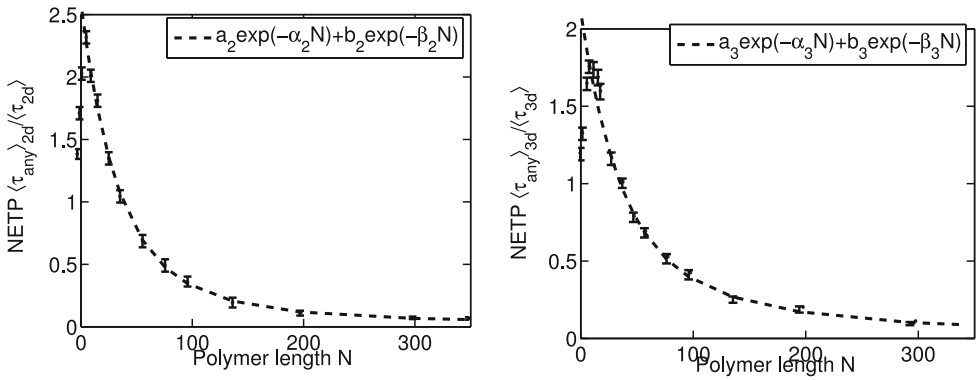


Fig. 21: **NETP of a monomer to the target for various polymer lengths.** Mean time that any one of the monomer to reach a small target located on the boundary (normalized to the NET τ_0 for a single monomer) in two and three dimensions **(a)** and **(b)**. Each point is an average over 2000 runs. Brownian simulations are fitted by a double exponential $\frac{\langle \tau_{\rm any}(N)\rangle_{2d}}{\langle \tau_{2d}\rangle} = a_2 \exp(-\alpha_2 N) + b_2 \exp(-\beta_2 N)$ with exponents $\alpha_2 = 0.0075, \beta_2 = 0.024$ and coefficients $a_2 = 0.23, b_2 = 2.17$ and in dimension 3 by $\frac{\langle \tau_{\rm any}(N)\rangle_{3d}}{\langle \tau_{3d}\rangle} = a_3 \exp(-\alpha_3 N) + b_3 \exp(-\beta_3 N)$ with exponents $\alpha_3 = 0.0082, \beta_3 = 0.030$ and coefficient $a_3 = 0.60, b_3 = 1.56$. (reproduced from [8])

The bell shape nature of the NETP can be qualitatively explained using the NET equations: indeed, a small polymer can be considered as a quasi-particle of radius $R_g = l_0\sqrt{N/6}$ [57], evolving in an effective domain which is the full domain minus its volume. Thus the NETP is related to the mean first passage time of the quasi-particle with diffusion constant $D_{\rm CM} = \frac{D}{N}$ in the apparent domain of volume $V_a = \frac{4\pi}{3}(R - R_g)^3$, leading to a mean time proportional

$$\frac{V_a}{\varepsilon D_N} = N\frac{4\pi}{3\varepsilon D}(R - l_0\sqrt{N/6})^3. \tag{168}$$

This phenomenological formula shows that the mean time has a maximum for $N_m = 25$, which is an over estimation of the empirical value that we obtained from Brownian simulation $N = 10$.

The case when only one monomer can find the target is quite different from the situation where they can all find it: indeed the monomer location along the polymer chain influences the result of the NETP, as shown in Fig. 22a. When a polymer model is confined in a ball and only one monomer can be absorbed and all others are reflected at the target site, numerical simulations reveal that

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between the middle and the end monomer, there is a factor 3 reduction in the arrival time (Fig. 22a). Interestingly, already taking the $N-1$th monomer compared to the last one is making a noticeable difference. In addition, the NETP increases to a different plateau that depends also on the monomer location. During the increasing phase, the arrival of the monomer to the target is mainly governed by the diffusion of the polymer center of mass, which can be approximated by a diffusing ball of diffusion constant $D_{\mathrm{CM}} = D/N$. However, for larger N, this approximation is not valid, rather the arrival time converges to a constant value that depends on the stochastic dynamics, which is the one of a correlated particle (monomer) to a target. Complementary results are discussed in [100].

Finally, we have shown in Fig. 22b, that the histogram of arrival time of any monomer to a small target can well be approximated by a single exponential, suggesting that the arrival time is almost Poissonian. This situation is similar to the arrival time of a single monomer to a small target located on the wall of a cavity, as shown in Fig. 19a and c. In both cases (search for a small target by a single monomer for for one end by the other end of a polymer in a confined domain), the NET converges asymptotically to a value in the large polymer length limit(Fig. 19b and d and Fig. 22a). The most striking difference is due to the apparent potential well generated by the reflections of the monomers on the boundary that increases the search time, which is not the case for Rouse polymers. But in both cases, the Poissonian statistics reflects that the searches for a small target is a rare event, characterized by long time asymptotics [239].

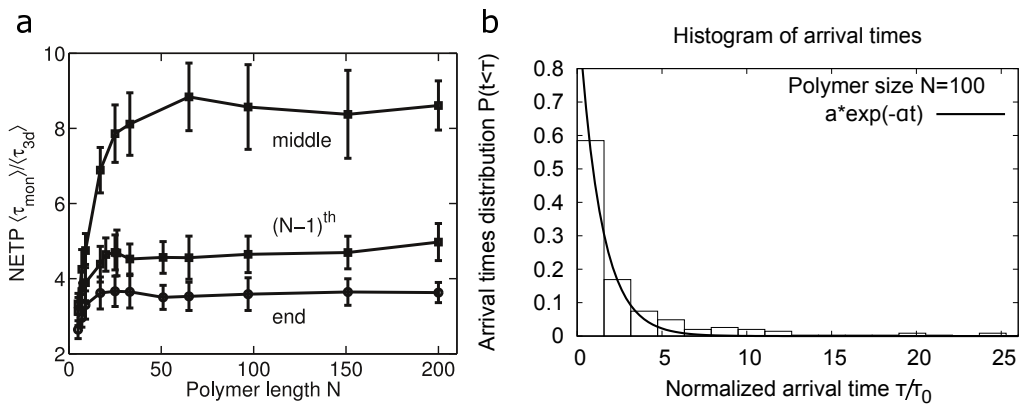


Fig. 22: **Mean search time of a single monomer (NETP) to a small hole located in a cavity (see also figure 3). (a)** NETP for three different monomers: end, middle and N-1: The NETP is plotted as a function of the polymer length N in three dimensions (Brownian simulations): The encounter time is normalized to the time τ_0 (for a single Brownian particle). Parameters are described in table 2. **(b)** Probability distribution $P[\tau/\tau_0]$ of arrival times for the end monomer to a small target (in three dimensions). The Probability distribution of the arrival times to a small hole located on the boundary of a sphere in three dimensions is well approximated by a single exponential $Pr\{\tau_{3d} = t\} = a\exp(-\lambda t)$ with $a = 1.014, \lambda = 0.76$ (reproduced from [8]).

4.13 Induced screening potential on a monomer by the other ones at a boundary

The interaction of a monomer with other monomers generates an effective potential through their interaction with the boundary, which is different for the middle and the end monomer, leading to the major differences in the searching time, as shown in Fig. 22a. For N large enough, the middle monomer is more confined than the end one, and by analogy with the diffusing of a stochastic particle

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in a potential well, the middle monomer has to surmount a higher potential barrier to reach the small target located on the boundary. Indeed, a crude approximation consists in considering that the active monomer motion with position $\boldsymbol{X}(t)$ follow a diffusion process in a spherical symmetrical potential V inside a confined spherical domain,

$$\dot{\boldsymbol{X}} = \frac{V(\boldsymbol{X})}{\gamma} + \sqrt{\frac{2kT}{\gamma}}\dot{\boldsymbol{w}}, \tag{169}$$

where $\boldsymbol{w}$ is the standard Brownian motion. Using the symmetry of the domain, the potential V has a single minimum at the center. In a high potential barrier approximation, the mean time to a small target does not depend on the specific shape of the potential, but rather on its minimum and maximum [248]. In that case, the mean time $\tau(N)$ to a target, which depends on the polymer size is given by:

$$\tau(N) = \frac{(2\pi)^{3/2}\gamma\sqrt{kT}}{4a\omega_N^{3/2}} \exp\left[\frac{U_N}{kT}\right] \tag{170}$$

where a is the size of the small target, $U_N(r)$ is the energy potential generated by the interaction between the polymer and the boundary of the domain at the target site. The key parameter is the frequency ω_N at the minimum of $U_N(r) \approx \frac{\omega_N}{2}r^2$ near 0 [248]. The potential $U_N(r)$ can be recovered from formula 170 and it is approximated by the numerical simulations described in Fig. 23c. Interestingly, depending on their position along the polymer chain, monomers have also different three dimensional spatial distribution inside a ball as revealed by numerical simulations (see Fig. 23a,b). For example, the middle monomer is more restricted to the center compared to the end monomer, which explores in average a larger area. When the length of the chain increases, the middle monomer position become more restricted, while the end-one does not seem to be much affected (Fig. 23a). The pdf of the center of mass position is comparable to the one of the middle monomer. Today, there are no analytical formula for the mean distance of a monomer to the boundary of a bounded. The extrusion of the end monomer from the center is probably a consequence of a single spring force at the end, compared to two for other polymers.

Using the simulated pdf, it is possible to define an effective potential $U_N(r)$ acting on a single monomer using the relation $U_N(r) = -k_BT\log(P(r))$ (derived from eq. 170). Interestingly, the potential acting on the center of mass is large enough, so that it cannot reach the periphery of a ball during simulations (Fig. 22c,d) [8]. A direct derivation of this potential is laking but would certainly be very useful to study the collective effect of many monomer on one monomer through the boundary.

Finally, similar to the case where all the monomers can find a small target, the arrival time distribution of a single monomer to a small hole is well approximated by a single exponential. To conclude, the distribution of arrival time of a monomer to a small target is almost Poissonian, however the rate depends on the location of the monomer along the polymer chain, that can be absorbed at the target site. NETP can also be studied when additional constraint such as stiffness is added to the model NETP (see for example [8]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

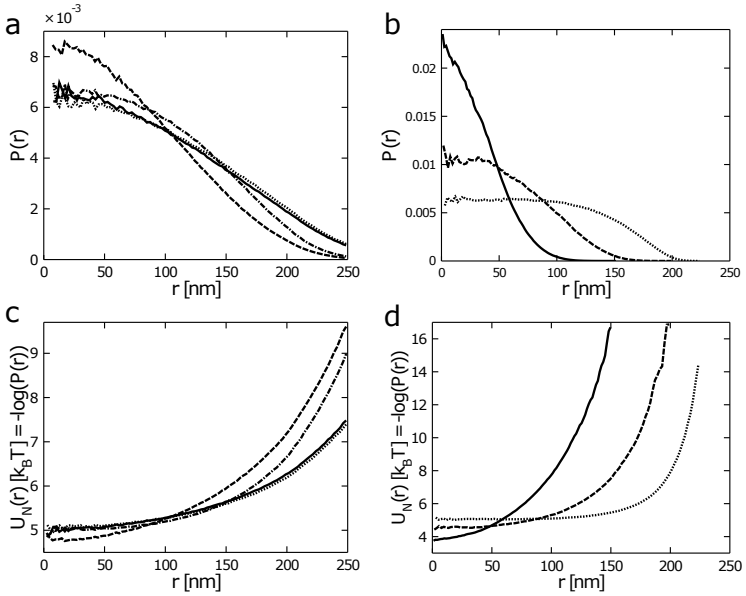


Fig. 23: **Distribution of monomer position and the influence of the boundary.** **(a)** Probability distribution function (pdf) of monomer positions. The radial pdf is computed for the end $N = 16$ (points), $N = 150$ (full line), the middle $N = 16$ (points-line) and $N = 150$ (dashed line) monomers. **(b)** The pdf for the center of mass with $N = 16$ (points), $N = 48$ (dashed line), $N = 150$ (full line). The normalized effective potential $U_N(r) = -k_BT \log(P(r))$ acting on a monomer is computed from the radial pdf. **(c)** The potential for the end and middle monomers is shown in Fig. 22a. **(d)** The potential energy for the center of mass. The behavior for the end and the middle monomers are significantly different, confirming that the boundary has different effect depending on the position of each monomer. The parameters are summarized in table 2 (reproduced from [8]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4.14 NETP with additional stiffness

Adding rigidity on the polymer can also influence NETP. Indeed chromatin stiffness can be modulated by nucleosomes or bound proteins. Stiffness is accounted for by including the bending energy

$$U_{bend}(\boldsymbol{x}) = \frac{\kappa_{ang}}{2}\sum_{i=1}^{N-1}(\boldsymbol{u}_{i+1}-\boldsymbol{u}_i)^2 = \kappa_{ang}\sum_{i=1}^{N-1}(1-\boldsymbol{u}_i\cdot\boldsymbol{u}_{i+1}), \tag{171}$$

as described in subsection 3.8. With parameter described in that subsection, stochastic simulations for the arrival time of any monomer to the absorbing boundary are shown in fig. 24, suggesting that the possible relation in dimension 2 (Fig. 24a)

$$\frac{\langle\tau_{\text{any}}\rangle_{2d}}{\langle\tau_{2d}\rangle} = a_2\exp(-\alpha_2 N) + b_2\exp(-\beta_2 N) \tag{172}$$

with $a_2 = 0.37, b_2 = 2.9$ and $\alpha_2 = 0.02, \beta_2 = 0.18$ and in dimension 3, (Fig. 24b)

$$\frac{\langle\tau_{\text{any}}\rangle_{3d}}{\langle\tau_{3d}\rangle} = a_3\exp(-\alpha_3 N) + b_3\exp(-\beta_3 N) \tag{173}$$

with $a_3 = 0.28, b_3 = 1.64$ and $\alpha_3 = 0.01, \beta_3 = 0.12$. Compared to the nonflexible polymer, the maximum of the NETP is now shifted towards smaller values of N: The NETP is an increasing function of N for $N < 6$, and for large N, it is a decreasing function of N.

In a confined spherical cavity, flexible polymers have a tendency to fill the available space and the probability of finding a monomer at the center of the cavity is higher compared to the boundary, when the persistence length is of the radius of the ball. On the contrary, for stiff polymers, the polymer chain has to bend abruptly near the boundary, where a large fraction of the polymer is found (Fig. 24c-d). Thus the search time for a target located on the surface by a monomer located on a nonflexible polymer is thus facilitated and the search should be almost two-dimensional, leading to a decreased NETP. To conclude, nonflexible polymer find small targets faster than completely flexible ones, this is due to an increase probability to find monomers in the close vicinity of the boundary where the small target is located.

4.15 Summary of the looping time formulas

We now summarize the asymptotic formula for the looping time (MFLT) in dimensions two and three respectively, for small ε.

4.15.1 Mean First Encounter Time formula in free space for a Rouse polymer

$$\langle\tau_\varepsilon\rangle_{2d} = \frac{N}{2D\kappa}\log\left(\frac{\sqrt{2}b}{\varepsilon}\right) + A_2\frac{b^2}{D}N^2 + \mathcal{O}(1), \tag{174}$$

$$\langle\tau_\varepsilon\rangle_{3d} = \left(\frac{N\pi}{\kappa}\right)^{3/2}\frac{\sqrt{2}}{D4\pi\varepsilon} + A_3\frac{b^2}{D}N^2 + \mathcal{O}(1), \tag{175}$$

where ε is the radius centered at one end, D is the diffusion coefficient, $\kappa = dk_BT/b^2$ is the spring constant with d the spatial dimension, k_B is the Boltzmann coefficient and T the temperature and

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

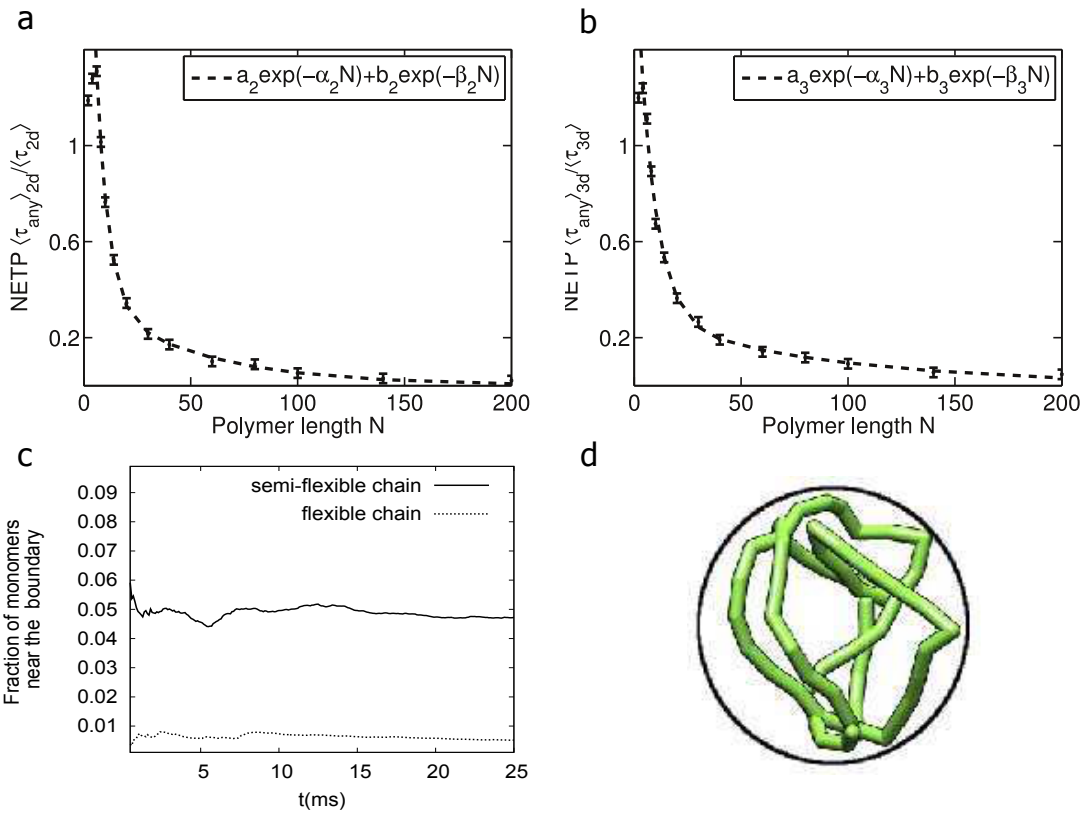


Fig. 24: **NETP for a semi-flexible polymer**. Numerical simulations are performed when any of the monomers (of a semiflexible polymer) can reach the small target in dimension 2 (a) and 3 (b). The NETP is normalized to τ_0 (the NET for 1 bead). Each point is an average over 2000. A double exponential fit $\frac{\langle\tilde{\tau}_2\rangle}{\langle\tau_{2d}\rangle} = a_2\exp(-\alpha_2 N) + b_2\exp(-\beta_2 N)$ leads to $a_2 = 0.37, b_2 = 2.9$ and $\alpha_2 = 0.02, \beta_2 = 0.18$ and in dimension 3, $\frac{\langle\tilde{\tau}_3\rangle}{\langle\tau_{3d}\rangle} = a_3\exp(-\alpha_3 N) + b_3\exp(-\beta_3 N)$. In three dimensions, the parameters are $a_3 = 0.28, b_3 = 1.64$ and $\alpha_3 = 0.01, \beta_3 = 0.12$. **(c)** Fraction of time a monomer spends near the boundary ($> 0.9R$, R is the radius) for a flexible and nonflexible polymer. **(d)** snapshot of a nonflexible polymer where a large fraction of its monomer are located near the spherical boundary (reproduced from [8]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A_2 and A_3 are constants, fitted to simulations data (see fig. 17 for explicit values). Moreover, the distribution of looping time is

$$p(t) = Pr\{\tau_\epsilon > t\} = \int_{\Omega_\epsilon} p(\boldsymbol{x}, t)dx = \sum_{i=0}^{\infty} C_i e^{-\lambda_i^\epsilon Dt} \tag{176}$$

where C_i are constants and is well approximated by a sum of two exponentials. For N ($N = 16$ and 32) not too large, a single exponential is sufficient,

$$p_N(t) \sim e^{-\lambda_N t} \tag{177}$$

with $\varepsilon = 0.1b$. Here $\lambda_{16} = 0.0125b^{-2}, \lambda_{32} = 0.0063b^{-2}$. For long polymers, a sum of two exponentials is more accurate

$$p_N(t) \sim C_0 e^{-\lambda_0^\epsilon t} + C_1 e^{-\lambda_1^\epsilon t}. \tag{178}$$

For $N = 64$, the numerical values are $\lambda_0^\epsilon = 0.0012b^{-2}, \lambda_1^\epsilon = 0.0375b^{-2}, C_0 = 0.99, C_1 = 0.28$. Although the two exponential approximation works well for small $\varepsilon < 0.2b$, four exponents are needed for larger ε ($> 0.4b$). For $N \in [4 - 64]$, $C_0 \approx 1$, while C_1 remains approximately constant for a given value of ε. For example, for $\varepsilon = 0.1b$, C_1 varied with N from 0.2 to 0.28.

4.15.2 Mean First Encounter Time formula in a confined domain for a Rouse polymer

The MFET in a confined ball of radius A is estimated when the boundary is accounted for by a parabolic potential, added to the Rouse potential ϕ_{Rouse} such that the total energy is the sum

$$\phi_{\text{h}} = \phi_{\text{Rouse}} + \frac{\beta}{2}\sum_{n=1}^{N} \boldsymbol{R}_n^2 = \frac{1}{2}\sum_{p=0}^{N-1}(\kappa_p + \beta)\boldsymbol{u}_p^2, \tag{179}$$

where $\kappa_p = 4\kappa \sin^2(\frac{p\pi}{2N})$, $\boldsymbol{u}_p$ are the coordinates in which ϕ_{Rouse} and the strength B is calibrated to the radius of the ball by

$$B = \frac{12}{A^4/b^2 + 2A^2}, \tag{180}$$

so that the root mean square end-to-end distance of the polymer in the potential field is equal to the square radius of the confining ball domain A, that is [14]

$$\langle(\boldsymbol{R}_N(B) - \boldsymbol{R}_1(B))^2\rangle = A^2 \quad \text{for } N \gg 1. \tag{181}$$

In that case, the MFET in a confined ball of radius A $\langle\tau_{\text{h}}\rangle$ for two end monomers of a Rouse polymer to meet is given for $\varepsilon \ll b$,

$$\langle\tau_{\text{h}}\rangle \approx \frac{2^{1/2}}{4\pi\varepsilon D}\left[\frac{4\pi N}{N^2 B + \pi^2\kappa} + \frac{4}{\sqrt{\kappa B}}\left[\frac{\pi}{2} - \tan^{-1}\left(2\sqrt{\kappa/B}\tan\left(\pi/2N\right)\right)\right]\right]^{3/2} + \mathcal{O}(1). \tag{182}$$

Note that as N tends to infinity, the mean looping time $\langle\tau_{\text{h}}\rangle$ does not diverge to infinity, but converges to an asymptotic value $\frac{1}{\pi\varepsilon D}(\frac{\pi}{\sqrt{\kappa B}})^{3/2}$. Finally, the distribution of looping times is always well approximated by a Poissonian law $p(t) \approx e^{-\lambda_0^\epsilon Dt}$ in contrast to looping in a free space.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4.15.3 Mean First Encounter Time formula for a β-polymer.

The mean encounter time for the two end monomers of a β-polymer is given by

$$\langle \tau_\epsilon^B \rangle \approx \frac{1}{D\varepsilon(2\kappa)^{3/2}} \left[\frac{2}{3} \cos\left(\frac{\pi}{2N}\right)^3 {}_2F_1\left(\frac{3}{2}, \frac{1+\beta}{2}, \frac{5}{2}, \cos\left(\frac{\pi}{2N}\right)^2\right) + \frac{\pi}{N} \frac{\cos^2\left(\frac{\pi}{2N}\right)}{\sin^\beta\left(\frac{\pi}{2N}\right)} + \frac{\pi^2}{12N^2} \frac{\beta\pi \cos\left(\frac{\pi}{2N}\right)^3}{2\sin(\frac{\pi}{2N})^{1+\beta}} \right]^{3/2} \quad (183)$$

and for large N, it is possible to use the estimation

$${}_2F_1\left(\frac{3}{2}, \frac{1+\beta}{2}, \frac{5}{2}, \cos\left(\frac{\pi}{2N}\right)^2\right) = 3\pi^{3/2-\beta} 2^{-(1+\beta)} N^{\beta-1} + \mathcal{O}(1) + \mathcal{O}(N^{-2}), \quad (184)$$

where ${}_2F_1$ is the Gaussian hypergeometric function [1]. The two other terms in 165 scale as $N^{\beta-1}$. This asymptotic result is quite different for the Rouse polymer which scales as $N^{3/2}$, while for a β-polymer, for $N \gg 1$, the MFET scales as $N^{\frac{3}{2}(\beta-1)}$.

4.15.4 Arrival time of a monomer to a small target located on the boundary of a bounded domain

Although, we are still missing an analytical derivation for the mean arrival time of a single monomer of a Rouse polymer to the boundary of a bounded domain, the approximation of this process as a single stochastic particle trapped in a single well lead to (see eq. 170)

$$\tau(N) = \frac{(2\pi)^{3/2}\gamma\sqrt{kT}}{4a\omega_N^{3/2}} \exp\left[\frac{U_N}{kT}\right] \quad (185)$$

where a is the size of the small target, $U_N(r)$ is the energy barrier, generated by the polymer due to the presence of the boundary of the domain at the target site and ω_N is the frequency at the minimum ($U_N(r) \approx \frac{\omega_N}{2} r^2$ near 0) [248, 8] and $\omega_N \sim N$. Additional research is expected to derive this statement from polymer model analysis.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5 Analyzing chromatin organization and dynamics using polymer model

This section is dedicated to large-scale simulations of chromosomes and telomeres [274, 291, 272, 124] based on self-avoiding polymer models. The output of these simulations is the exploration of the polymer configuration space, the genesis of statistics that agree with experimental data. We shall see here that polymer models can reproduce chromosomal organization, observed in live cell imaging data in various species such as yeast or mammalian cells. We insist in subsection 6 on local chromatin reconstruction for the X-chromosome inactivation in female mouse embryonic stem cells using sophisticated polymer models.

Polymer looping remains the central event of chromatin organization and in parallel of presenting large scale simulations, we discuss the role of asymptotic analysis presented in previous sections 4 to interpret and extract information from empirical data. Asymptotic formula facilitated even further compared to simulations the exploration of the large parameter space. Indeed, the polymer parameter space remains too complicated to be continuously described by simulations. These simulations allow exploring and sampling small subregions of the configuration space, which is largely insufficient to classify structures, because the underlying spatial dimensions of the polymer space is very high and robust structures require high-dimensional geometrical parametrization. At this stage, simulations and asymptotic analysis provide complementary tools in searching for features in data. Interestingly, passage time formula between two sites reveals some key geometrical polymer features, without examining separately the topology or the possible polymer configurations.

5.1 Numerical simulations and chromatin dynamics

Polymer simulations are used to describe DNA and chromatin folding inside the interphase nucleus of eukaryotic cells, which usually involves multiple length scales (from few to hundreds nanometers). To access higher-order structure, folding in confined nuclear space or distinct territories, polymer models reveal how chromatin looping is used in transcriptional regulation and how chromatin organization mediates long-range interactions [109, 30]. Transient loop formation can originated from thermal fluctuations. Looping interactions that do not directly involve an enhancer-promoter pair can modulate their interactions, as shown in Fig. 25 [74]: the simulation results show a $3-5$ fold facilitation of enhancer promoter (E-P) contact frequency, comparable to observed changes in gene expression. In a chromatin model [74] where monomers are connected by harmonic bonds, with a 15 nm diameter representing 500 bps (approximately three nucleosomes), a permanent loop is formed when two monomers are connected with a harmonic bond of the same strength. Two such loops are formed in the simulations and a bending energy is imposed to account for the rigidity of the chromatin fiber. Monomers interact via a Lennard-Jones potential (purely-repulsive potential, truncated at an energy $U = 3kT$). Polymers are confined to a ball and initialized from an un entangled conformation. The probability distribution of loops for random walk and self-avoiding random walk was studied for chain lengths of size N =64, 128, 256 and 512, using Monte-Carlo simulations on a cubic lattice (boxes were of width L 64) and a density of r $12:5\%$, which is similar to the conditions of interphase nuclei [30]. With a total of 4096 monomers, chromosomes are initially equilibrated as self-avoiding walks. After an initial equilibration step, the Monte-Carlo algorithm samples different loop configurations. A loop is formed with a certain probability P when two monomers co-localized. The life time of a loop is Poissonian. Results of simulations are shown in Fig 26. The contact map is similar to one obtained by 5C data, for a polymer with different looping probabilities: in the self-avoiding walk polymer model (Fig. 26A), there are only few contacts between beads located far

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

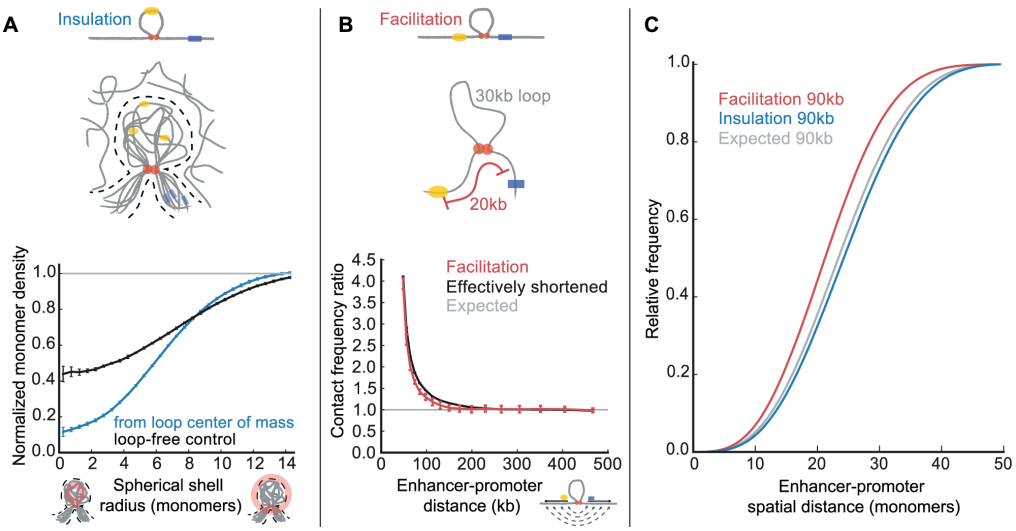


Fig. 25: **Mechanisms of insulation and facilitation.** (A) (top) Insulation mechanism: steric exclusion by a chromatin loop from the superposition multiple loops(grey, with enhancer and promoter) and their sterically excluded region (dashed lines), surrounded by other distal regions of chromatin (grey). (bottom) Density of distal monomers (i.e. outside the loop and > 10 kb from the loop base) as a function of radial distance from the center of mass of the loop. The loop-free control exactly repeats this procedure for an equivalent region without a loop. Both are normalized using respective radial-position dependent spatial density. (B) (top) Facilitation mechanism: an E-P pair flanking a loop has an effectively shorter genomic distance; here an E-P pair with 50 kb separation and a 30 kb loop behaves similarly to an E-P pair separated by 20 kb in a region without a loop. (bottom) Comparison of contact frequency ratios for the above situations, as a function of E-P distance. (C) Simulated cumulative distribution of spatial distances for an E-P pair with a genomic distance of 90 kb (reproduced from [74]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

apart. Increasing the looping probability (Fig. 26B and C) results in a strong increase of both the number of loops and the abundance of large loops [30].

The role of looping in large-scale (super Mega base pairs) folding of human chromosomes was stud-

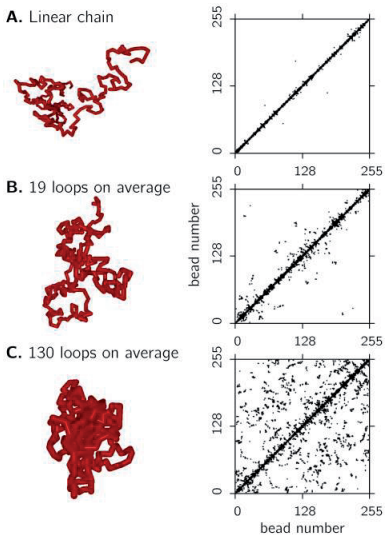


Fig. 26: **Intra-chromosomal contacts of isolated model polymers.** Equilibrated polymers with different looping probabilities. (A. linear chains (no loops), B. on average 19 loops per conformation and C. on average 130 loops per conformation) co-localized beads were determined and marked by a black square. For each image, the contacts of 4 independent polymer conformations are plotted. Linear chains (A) have not so many contacts between beads which are widely separated along the contour of the polymer. Increasing the probability of functional loops (B and C) results in a boost of contacts both between close-by segments as well as between segments having a large genomic separation (reproduced from [30]).

ied using self-avoiding walk polymer model that can generate transient looping [268]. Chromatin compaction was shown to relate to a reduction of the concentration of two looping proteins (knocked down the gene that codes for CTCF and the one coding for Rad21, an essential subunit of cohesin). This effect can be explained by selectively decreasing the number of short-range loops, leaving long-range looping unchanged [268]. Polymer configurations for different looping probabilities at short and long-ranges are shown in Fig. 27. Higher order loops involving more than two monomers was considered to model discrete spatial transcription foci: the formation of these foci resulted from the encounter of sparse interacting sites located on the same chromosome [135]. In these foci, several genes related to the same transcription system, come physically in close proximity in order to share regulatory and transcription proteins.

In summary, polymer models with additional short and long-range looping can account for chromatin organization at various scales and can be used to explain some of the modulation of gene activation.

5.2 Characterizing chromosomes and nucleus architecture using large scales simulations

Chromosomal territories are nuclear regions occupied by a single chromosome and are found in most higher eukaryotes, observed in live cell imaging data [32]. But lower eukaryotes, such as the yeast

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

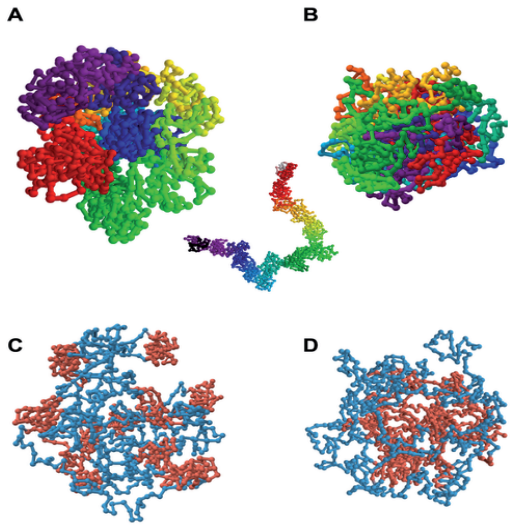


Fig. 27: **Polymer conformations of model at different looping regimes. [268].** (A) Conformation with high short-range (pshort = 0.12) and low long-range (plong = 0.04) looping probabilities and (B) the same polymer with low short-range (pshort = 0.04) and high long-range (plong = 0.12) looping probabilities. The color code labels the monomers of the polymer according to the visible spectrum along the length of the polymer. The inset in panel A displays the same situation as in (A) after abolishing all long-range looping (plong = 0), showing that a uniform thick fiber is formed. (C) Conformation of a polymer with high short-range (pshort = 0.16) and low long-range (plong = 0.02) looping probabilities and (D) the same polymer with low short-range (pshort = 0.02) and high long-range (plong = 0.16) looping probabilities. Here, topological domains are labeled red and non-looping linker regions blue (reproduced from [30]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Saccharomyces Cerevisiae, lack these chromosome territories and their chromosomes seem to be loosely arranged. Interestingly, in lower eukaryotes such as plants and flies, chromosomes are often polarized, with the ends of the arms (telomeres) on one side of the cell nucleus. The two arms are anchored at a single point (the centromere) located on the opposite side of the nuclear surface.

These structures are associated with gene territories, where genes of similar metabolic function are together [25]. Territories are revealed by microscopy imaging and also chromosome capture experiments, and are reproduced using large scales simulations [274, 291, 272, 212] of self-avoiding polymers. However, the output of these simulations remains difficult to compare to experimental data except at equilibrium, where the probability distribution function of chromosomes and moments can be estimated. Detailed simulations of higher eukaryotes are still lacking due to the complexity of the simulations, the number of degrees of freedom and the lack in understanding of the basic molecular mechanisms involved in establishing and maintaining these territories.

Transient re-organization of the genome is also hard to simulate, because it is difficult to identify relevant time scales. Although different coarse-graining simulations of the nucleus have been developed [83, 290], a universal model for chromatin is still to be found because it is difficult to identify steady or quasi-steady polymer states from experimental data (such as Hi-C). By taking into account constraints acting on the chromatin fiber, polymer simulations can reproduce measured physical quantities that scale with distances between genomic loci. Polymer models can also reproduce chromosomal territories, and probabilities of contacts between loci measured by chromosome conformation capture methods [83].

Simplified polymer model lead to various striking conclusions: classical models have ignored basic DNA identity such as sequences, but they accounted for the centromere, telomeres, and the ribosomal DNA (rDNA) to model the interphase yeast nucleus chromosomes. The first output is that the large-scale architecture of the yeast nucleus is dominated by random forces moving polymers, rather than by a large number of DNA-specific deterministic factors [290]. Recently, chromosomes were modeled as semi-flexible polymers with freely jointed chains of non-intersecting rigid segments (Fig. 28). This coarse-grained model neglected chemical bound details of the chromosomes, DNA sequence and possible histone modifications, but could reproduce chromosomal organization.

Using freely-joint-polymer chains requires specifying the following parameters: the length of each segment, equals to $50nm$ accounting for the chain rigidity, the number of base pairs corresponding to each segment (DNA compaction equal to 5 kbs), and the segment diameter (20 nm), see Fig. 28. Other values used in modeling and simulations are possible (persistence length of dsDNA is chosen around 50 nm and for chromatin fiber it is 150 nm [106]). A second output is that such model can replicate many of the steady-state statistics measured by contact frequencies across the genome (genome-wide chromosome-conformation capture), showing that the main statistical organization of chromosomes in the nucleus results from elementary stochastic polymer dynamics. This model was used to examine the impact of reducing the volume of the nucleolus: the model predicted a displacement of telomeres further away from the spindle pole body, in agreement with prior observations [269].

5.3 Interpreting chromosome capture data inside nuclear territories from polymer looping

Looping properties of polymer model provide constraints on the statistics of long-range interactions and the anomalous exponent that can be extracted from chromosome capture (CC) data. Indeed, analytical expressions for the MFET predict a decay exponent presented in eq.165. For example, long-range interactions introduced in the β-polymer model can be due to the condensin proteins,

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

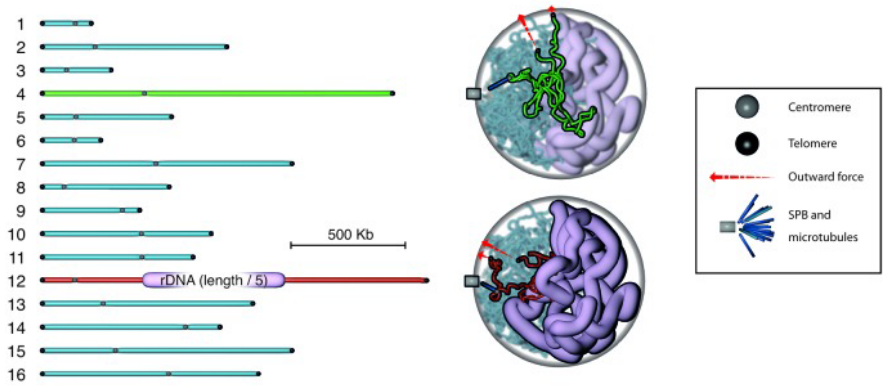


Fig. 28: **Computational model of chromosome architecture in the budding yeast nucleus [290].** The 16 chromosomes of the haploid yeast genome (shown stretched out on the left), are modeled as random polymer chains undergoing Brownian motion and confined to a spherical nucleus (right). Only three DNA sequence-specific constraints are included: centromeres, which are attached to the SPB via short rigid microtubules; telomeres, which are tethered to the nuclear envelope by an outward force; and the rDNA locus, which is modeled as a chain of increased diameter. Chromosomes 4 and 12 are highlighted in the 3D snapshot of the simulation shown on the right (reproduced from [290]).

part of the Structural Maintenance of Chromosomes (SMC) protein family [292]. These proteins are capable of generating large loops that hold together sites located far apart along a chromosome. During mitosis, the concentration of condensin molecules increases, resulting in an increase of the Young modulus of the chromosome [169].

We now summarize the CC-data. They contains statistics of contact frequencies that require a physical model for interpretation. In a reconstruction procedure, sites that have high encounter frequencies should be positioned in closer proximity inside the nucleus domain. Given a set of encounter probabilities $P(x_1, x_2)$, $P(x_1, x_3)$, .. $P(x_1, x_N)$, $P(x_2, x_3)$,...,$P(x_N, x_N)$, where $x_1, x_2, ..., x_N$ are genomic sites, the following reverse engineering problem should be resolved: how to reconstruct the spatial positions $x_1, ..x_N$ from encounter probabilities? Is this reconstruction possible? This construction remains an open problem and requires additional information.

When sites are located on the same chromosome chain, using a Rouse polymer, the distances between any site in a free space increase with the distance $d(x_n, x_m) = \sqrt{\langle (x_m - x_n)^2 \rangle} \sim |m - n|^{1/2}$, while the encounter probability decays with the genomic distance as

$$p(x_n, x_m) \sim |n - m|^{-3/2}. \tag{186}$$

When sites are not located on the same chromosome, the encounter probability is a characteristic feature of the long-range nuclear organization and a reconstruction procedure should account for this decay. We should assume that the genome is at equilibrium. By averaging over millions of cells, the CC-data is averaged over population [152]. In reality, the chromatin is never at equilibrium, as it is constantly pulled by different proteins and sometimes open during transcription. In this case, interpretation of the data should take into account the dynamics and the encounter probability given by

$$P_{\text{enc}}(x_n, x_m) = \frac{\langle \tau_\epsilon \rangle^{-1}(x_n, x_m)}{\sum_{l=l_{\min}}^{l_{\max}} \langle \tau_\epsilon \rangle^{-1}(x_n, x_l)}, \tag{187}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where $\langle \tau_\epsilon \rangle^{-1}(x_n, x_m)$ is the rate of encounter between sites x_n and x_m [9].

To conclude, a procedure for reconstructing the chromatin organization should assume that polymer equilibrium is reached. So far, there are no generic procedures to reconstruct chromatin organization from data. As we shall see in the next sections, not only a polymer model should be chosen, but also a spatial scale (distance between two monomers) and some constraints on the empirical statistics to be prescribed. For example, it is possible to construct a polymer model such that the encounter probability of the simulations and the empirical data is the same [246]. Other criteria may be chosen as well. These constraints insure that the reconstructed model contains some of the key statistical characteristic of the data.

5.4 Interpreting CC-data using polymer models

At a resolution of 1Mbp, chromosomes can be divided into segments, with higher inter-segment contact frequencies between different segments [157]. These sub-segments form blocks in the frequency matrix. The encounter probability decays as a power-law of the genomic distance ($P_{\text{enc}} \sim n^{-1}$) (see fig.29a). An exponent decay of 1.5 suggests an equilibrium polymer and corresponds to a flexible polymer model. However, the power law extracted from data differs from 1.5, leading to a constraint for the choice of polymer models. Various models have been proposed to explain qualitatively these observations, as we shall now review.

5.4.1 Fractal polymer globule

The compact conformation frequency data suggests that the polymer conformation of chromosomes followed a "fractal globule" image [96, 97]: this non-equilibrium state polymer consists of non-entangled structure, where beads are crumpled in small non-overlapping domains, which gradually fill a domain according to a auto-similar rule, without crossing [183] (Fig. 29b). The DNA may be compacted in a non-entangled manner in the nucleus, such that its opening during transcription would be not impaired by any nodes. This representation is now clearly insufficient to account for the chromatin organization [196].

5.4.2 Interpreting CC-data using the $\beta-$polymer model

The spatial organization of the chromatin can also result from specific protein interactions. In the context of the β-polymer model, the encounter probability P_{enc} can be computed from relations 154 and 155. The normalized sums 159 and 160 for the encounter probability

$$P_{\text{enc}}(N) = \frac{\langle \tau_\epsilon^\beta \rangle^{-1}(N)}{\sum_{N'=N_{\min}}^{N_{\max}} \langle \tau_\epsilon^\beta \rangle^{-1}(N')}, \tag{188}$$

reveal the decay rate shown in Fig. 29c for $N_{\min} = 4$ and $N_{\max} = 80$: the encounter probability P_{enc} does not necessarily decay with $N^{-3/2}$ nor N^{-1}, but $N^{-3/2(\beta-1)}$, where β is identified by fitting a curve to data [157].

To conclude, it is possible to relate the empirical decay rate to the intrinsic properties of the chromatin, incorporated in the value of the parameter β.

5.4.3 String binder and switch model (SBS) model

Although chromatin is quite compacted, another approach to interpret the CC data is provided by the string-binder and switch model [19, 20] (Fig. 29d), where loops are generated randomly inside

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

a ball. The chromatin is constraint with many binding sites that represent different DNA binding proteins [171]. To account for protein interactions, chromatin is represented as a flexible chain with self-avoiding interactions (Fig. 29d) [194, 195]. A fraction f of the polymer sites can bind particles (binders) with a binding energy E_X. The binder molecules can bind to more than one monomer, allowing transient connections to be formed between polymeric sites. These binders can represent chromatin molecules such as the CTCF Zinc finger protein [136], known to mediate inter- and intra-chromosomal contacts [293].

Polymer conformation were studied using Monte-Carlo simulations for a concentration c_m of binder molecules. Adding binders affects the polymer size and can lead to collapse at a specific binder concentration C_{tr}. Interestingly, changing the concentration of the binders can reproduce long-encounter in chromosome capture experiments (Fig. 29e).

In summary, the SBS model accounts for the compaction observed in CC data based on specific interaction acting between distance monomers. However, the exact location and numbers of binders discussed in [194, 195] is not yet constraint and the empirical encounter probability cannot be recovered. Interestingly, it should be possible to account for these binding interactions that can further change during development or transcription.

With the increasing richness extracted from CC data, their interpretation has moved from general mean field models (fractal globule, β-model, SBS) to more precise but heterogeneous polymer models. This change allows exploring the reorganization of mitotic Chromosome [192] or the relationship between transcription and chromosome conformation of the X-inactivation center, where the regulatory element are located on the same chain (*cis* regulation [91]), see below.

5.5 Using CC-data and encounter probabilities to describe nuclear sub-confinement

The encounter probabilities (relation 188) between chromosomal sites at steady state is used to extract the confinement radius from measured looping distribution obtained from 4C data of yeast *Saccharomyces cerevisiae* [75]. Indeed, the frequency of encounter is inversely proportional to the MFETC, thus the encounter frequency can be obtained from the MFETC formula (eq. 165). Although the MFETC can vary with the monomer position along the polymer chain that it will encounter, the normalized rate (the reciprocal of the MFETC divided by its integral over the length) of monomer interaction (as experimentally measured) is independent on the encounter rate radius ε (as it cancels out in the normalization). The encounter probability depends only on the distance between two monomer pairs along the chain and is in fact not much affected by the remaining polymer tails.

The chromatin molecule can be organized in higher order structure such as stable separated domains that result from local specific interactions. At the current stage, the analysis of the model does not take into account the effects due to higher order structure in chromosomes. Another limitation of this model when applied to yeast, is due to the Rabl chromosomal configuration where the centromeres are connected to the mitotic spindle body. This interaction breaks the radial symmetry. The above approach was applied to 4C data in yeast, neglecting higher order structures and considering that the radius of yeast nucleus is fairly small (1μm).

The encounter probability data was computed from data presented in [75] and for a locus at position 99kbp from the right-end of chromosome II in the yeast (Fig. 30d) and fitted with the encounter formula eq.128, where β is the only free parameter, leading to $\beta = 3.91 \times 10^{-6}\text{nm}^{-2}$. Using formula 180, an effective confinement radius was extracted [9] of $A = 230$nm, representing a subdomain of the yeast nucleus [269]. Moreover, for large distances along the strand, the encounter probability reaches an asymptotic value rather than going asymptotically to zero (Fig. 30d). This decays shows

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

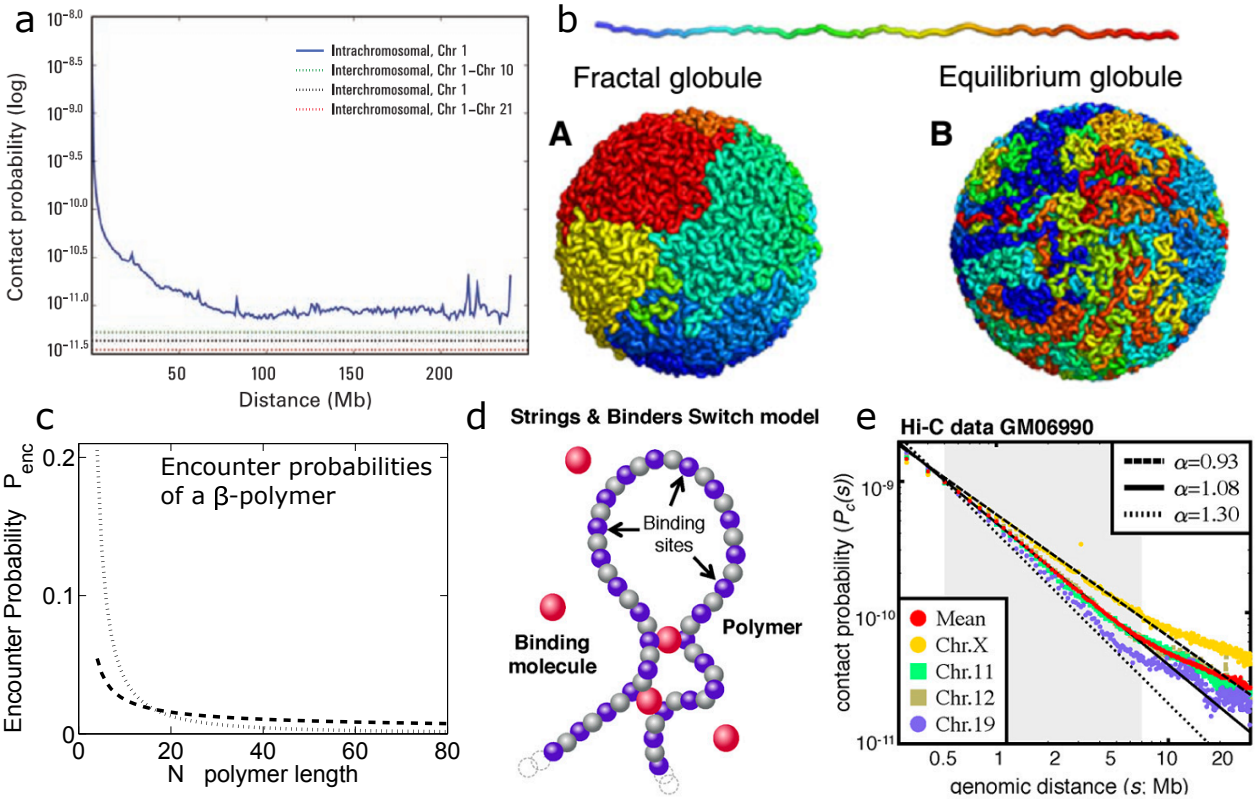


Fig. 29: **Encounter frequencies of genomic sites inspired chromosomes polymer models:** (**a**) Probability of contact decreases as a function of genomic distance on chromosome 1, eventually reaching a plateau at 90 Mb (blue). The level of interchromosomal contact (black dashes) differs for different pairs of chromosomes; loci on chromosome 1 are most likely to interact with loci on chromosome 10 (green dashes) and least likely to interact with loci on chromosome 21 (red dashes). Interchromosomal interactions are depleted relative to intrachromosomal interactions. [157] (**b**) Conformations of the fractal (A) and equilibrium (B) globules. The chain is colored from red to blue in rainbow colors as shown on the top. The polymer globule has a striking territorial organization, which strongly contrasts with the mixing observed in the equilibrium globule. (**c**) The normalized encounter probability for the polymer ends for the Rouse polymer for which $\beta = 2$ (points, $P(N) \sim N^{-1.5}$) and a β polymer for which $\beta = 1.5$ (dashed line, $P(N) \sim N^{-1.5}$) [10]. (**d**) Schematic of the string binder and switch model. The chromatin is modeled as a SAW polymer [19]. A fraction of the monomers (purple) can bind binder molecules (magenta) diffusing in the solution, which transiently fixed polymer looping events. (**e**) Contact probability, $P_c(s)$, was calculated separately for different chromosomes from published Hi-C dataset in human lymphoblastoid cell line GM06990 [157]. Chromosomes 11 and 12 follow the average behavior reported (reproduced from [157]) in the 0.5-7 Mb region (shaded in grey), with exponent α of approximately 1.08. Chromosomes X and 19 deviate from the average, with α exponents of approximately 0.93 to approximately 1.30, respectively. In a given system, different chromosomes can have different exponents.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

that the encounter of loci pairs is affected by nuclear confinement. Extending the analysis of polymer models, beyond the Rouse linear chain would be relevant to account for higher-order organization of the chromosome. Geometrical constraints of chromatin loci are classically obtained by single particle tracking of a florescent loci on the chromatin. Extracting spatial information from CC data, which can be later on compared to single particle tracking experiments would certainly lead to novel results about the organization of local subnuclear compartments.

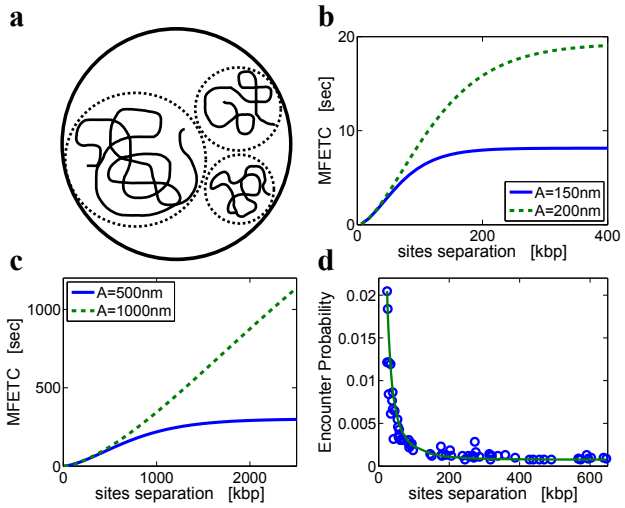


Fig. 30: **Encounter between two DNA loci inside a domain**. (**a**) Schematics representation of gene territories inside the nucleus. (**b-c**) MFETC of two sites located on the same chromosome as a function of their separation along the DNA strand. Chromosomes are confined in domains of radii $A = 150, 200, 500, 1000$nm. The encounter radius is $\varepsilon = 5$nm. (**d**) Encounter probability of a locus on chromosome II in the yeast and fitting of our looping probability formula. The effective confining radius extracted using our calibration formula is $A = 240$nm. Parameters are $b = 30nm$ and $D = 310^3 nm^2 s^{-1}$ (reproduced from [9]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6 Chromatin modeling using polymer

Polymer models are used to extract specific features from chromosomal capture data, presented in section 5. There is a large diversity of genome folding. For example, metazoan genomes are folded in a hierarchy of structures, while in mammals, each chromosome is folded independently of each other and occupies a well-defined territory (in interphase). The ability of higher Eukaryotes to regulate and differentiate depends precisely on different chromatin states. Chromatin domains vary in length between kilo to Megabps, but are separated by boundary-imposing-elements called insulators. Eukaryotes have chromatin domains with distinct functionality. The open chromatin called the euchromatin less dense than the rest permits a higher accessibility of proteins to gene sites. However, the dense heterochromatin region, much less accessible to transcription machinery is associated with gene silencing.

Inside each chromosome territory, some sub-compartments have emerged, corresponding to large (multi-Mbps) active and inactive chromosomal regions, which tend to interact with regions of similar transcriptional activity and are segregated from each other [157]. Interestingly, these compartments were found to be cell-type and developmental-stage specific, as they reflect the transcriptional state of a chromosome. Recently, sub-compartment folding units called topologically associating domains (TADs) or topological domains [71, 197, 121, 245] were characterized as sub-megabases (around 800kbs) domains with increased three-dimensional interactions, imposing a genome partition in adjacent regions of preferential chromatin contacts. TADs are further partitioned into sub-TAD domains and loops near binding factor (CTCF) sites [214, 253, 280] and co-factor cohesin [20]. Gene expression depends on distal regulatory regions called enhancers, which can be located mega bps away from the controlled gene [240] and TADs regulate the interactions of distal regulatory elements and enhancers [164, 260].

Understanding how TADs are established and maintained, how is the three-dimensional conformation of chromatin within single TADs the mechanisms has benefited from polymer models. Indeed, mean-field approaches described in section 5 have been used to reconstruct the average three-dimensional conformations of genomic regions starting from 3C-based datasets [22, 75]. While these approaches can be useful to represent the conformation of a genomic region, they do not provide any mechanism to explain neither conformational changes nor can they account for the cell-to-cell variability. The statistical and dynamical properties of the chromatin fiber within single TADs should depend on genomic elements/DNA binding factors. To account for cell-to-cell variability and population heterogeneity, various classes of phenomenological polymer models were proposed [91, 270, 132, 246] and the goal of this section is to summarize them.

6.1 Chromatin modeled by co-polymer

Starting with a semi-flexible chains and self-avoiding interactions (see section 3.7), the configuration of a co-polymer is by definition driven by thermal motion and repulsive interactions. The resulting structure is a homogeneous chain, except at the polymer ends. Heterogeneous structures are generated by adding interactions in the energy term

$$\begin{aligned} H &= U + H_{\text{Inter}}, \\ H_{\text{Inter}} &= \sum_{n<m} E_{nm} \exp\left(\frac{r_{nm}^2}{2r_0^2}\right), \end{aligned} \tag{189}$$

where the potential U includes spring forces and LJ-interactions (eq.32), while the energy H_{Inter} accounts for specific interactions between different chromatin sites. The Gaussian potential H_{Inter}

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

depends on interaction strength E_{nm} between monomers n and m and r_0 and defines the range of interactions [132] (see also [274] for a Gaussian potentials).

To reproduce chromatin organization [44], Langevin dynamics simulations and Monte Carlo methods [198] are used to explore the configuration space [132]. Monomers with the same states interact directly following eqs.189 (Fig.31a). In that case, the interaction energy is the sum of two terms: $E_{mn} = U_{\rm ns} + \delta_{mn} U_s$, where the energy $U_{\rm ns}$ between every pairs of monomers accounts for the confinement into the nucleus and U_s is an attractive interaction between monomers having the same (epigenetic) state. Results of simulations show immiscibility between monomers of different types and phase separation of the polymer. This model was tested on region of Drosophila melanogaster chromatin [79], where five typical epigenomic states can co-exist: two euchromatic states and three heterochromatic states. The experimental Hi-C map (Fig.31bA) shows the internal folding (black) and active domains with almost uniform intra-domain contact probabilities, where inter-domain contacts are especially numerous (between black domains). A region is modeled by a chain of 131 beads with 10 kb per bead. The encounter probability is shown in Fig.31bB: the pattern of inter-black domain interactions and some long-range contacts between active domains can also be reproduced (Fig.31bC). The polymer section (black chromatin) forms a compact metastable globule that transiently dissociates, and small active domains are expelled at the periphery (Fig.31bD). While the model could account for some experimental properties, the initial hypothesis that different epigenomic states attract each other remains to be validated.

6.2 Chromatin condensation/aggregation depending on chromatin state

Chromatin de-condensation was recently visualized in Drosophila Kc167 cells [227, 28] revealing multiple genomic domains. These domains are divided into three major epigenetic states: transcriptionally active, inactive, and polycomb-repressed. This classification is made on the basis of histone modification and regulatory protein enrichment (Fig. 32a).

Active domains scaled differently in size than the repressed/inactive ones (see Fig.32b) and fitting a power-law for the relation between volume V and domain length L leads to $V \sim L^b$ where b is largest for the active state, decreasing for inactive and is the smallest for silence domains (Fig.32c,d). This result suggests that packaging density controls genetic activity. While active and inactive chromatin had a self-similar structure, which is a uniform power-law across large genomic length, the radius of gyration computed inside subdomain of the imaged chromatin behaves differently (see Fig.32): repressed chromatin is composed of smaller sub-domains with a $60\% - 80\%$ overlap.

6.3 Constructing an effective polymer model from Hi-C maps to sample the space of chromatin configuration

Chromosomal long-range interactions with an effective potential was recently used to describe chromatin dynamics [91]: this model permits extracting single-cell information from population-averaged 5C or Hi-C maps. Despite its descriptive power, as we have shown in a previous section, 5C and Hi-C are population-averaged assays in which the contact probabilities is averaged over populations of millions of cells. These data do not permit access to the cell-to-cell and temporal variabilities of the chromatin conformation. However, these data can be used to constrain polymer simulations based on the contact frequencies measured experimentally. Two key regulators of X-chromosome inactivation in female mouse embryonic stem cells [207] were analyzed using polymer simulations and used to reconstruct TADs, identified in the X chromosome inactivation system (Fig.33A). The promoters of Tsix and Xist are shown in Fig.33B.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

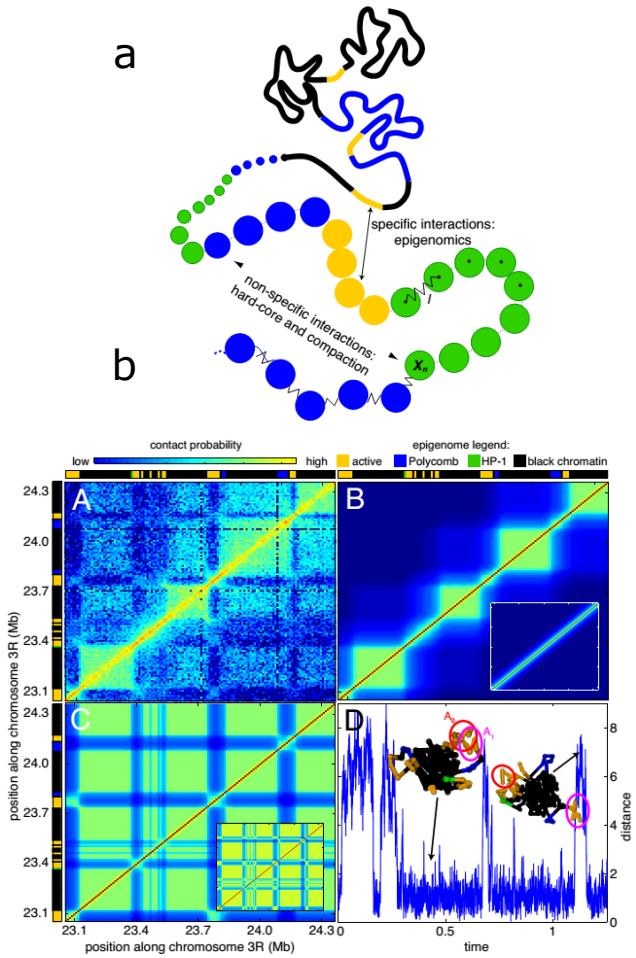


Fig. 31: **Chromatin modeled as a co-polymer**. (**a**) Block copolymer model: the chromatin is modeled as a self-avoiding bead-spring chain where each monomer represents a portion of DNA (10 kb) and is characterized by an epigenetic state: yellow (active), green (HP1-like heterochromatin), blue (Polycomb-like heterochromatin), black (repressive chromatin) [79]. The model integrates non-specific and specific short-range interactions to account respectively for the effective compaction of the chain and for epigenomically related affinities between monomers [132]. (**b**) (A) Experimental Hi-C contact map for the chromatin region located between 23.05 and 24.36 Mb of chromosome 3R (from [245]). Domains [79] represented at top and left borders are active (orange), Polycomb (blue), HP-1 (green) and black chromatin. (B and C) predicted contact maps inside the multistability region ($U_{ns} = -25$kBT, $U_s = -63$kBT) starting from a coil (B) or microphase separated (C) configuration (see insets). (D) Evolution of the distance between the centers of masses of the active domains A0 and A1 along a simulated trajectory. Insets: conformations of the chain (reproduced from [79]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

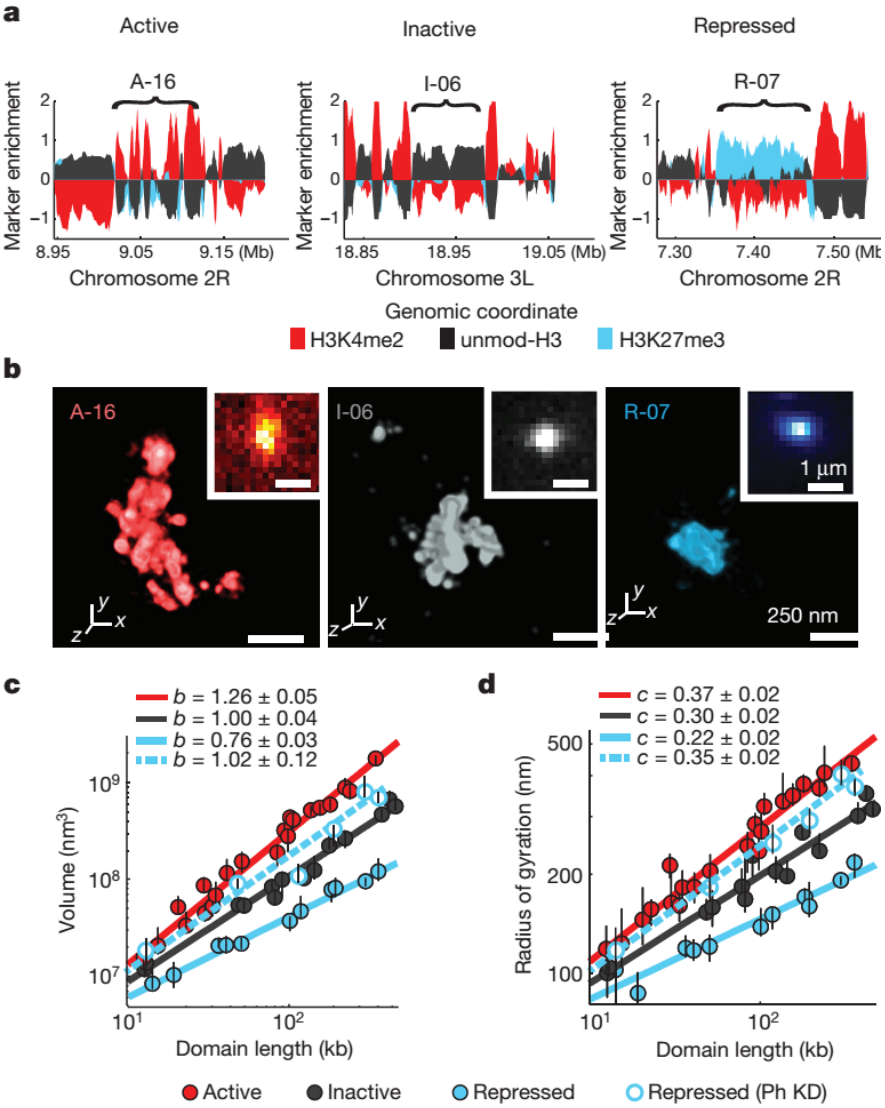


Fig. 32: **Chromatin in different epigenetic states exhibits distinct packaging and power-law scaling. (a)** Enrichment profile of H3K4me2 (red), H3K27me3 (light blue) and unmodified H3 (black) in three genomic regions, each harbouring an example Active, Inactive or Repressed domain (indicated by brackets). Marker enrichment, as defined in Supplementary Methods, was determined from ChIP-seq data20. **(b)** 3D-STORM images of three distinct epigenetic domains in **a**, labeled by in situ hybridization with DNA probes conjugated to the photoswitchable dye Alexa-647, shown with their corresponding conventional images in the inset. Each epigenetic domain appears as a single region in nearly all cells due to homologous pairing in the tetraploid Kc167 cells. **(c)** loglog plot of the median domain volume as a function of domain length for Active (red solid circles), Inactive (black solid circles) and Repressed (light blue solid circles) domains, as well as for Repressed domains in Ph-knockdown (Ph KD) cells (light blue hollow circles). Error bars represent 95% resampling ($n \approx 50$ cells). The lines indicate power-law fits, with the scaling exponent b shown in the legend. **(d)** As in **c** but the radius of gyration depends on the domain length with the scaling exponent c.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In this model, the chromatin fiber is coarse-grained as a chain of identical beads (108 for the Tsix TAD and 199 for the Xist TAD), separated by distance $a = 53$nm, which represents 3kb of genomic sequence (Fig.33A) [270]. This length is the average size of HindIII restriction fragments in the 5C data set in [197]. The original 5C data, based on pairs of interacting forward/reverse restriction fragments, is converted into a list of interacting bead-pairs sequences. The chromatin fiber is assumed to be at equilibrium (at least locally on the genomic length scale of a TAD), so that the probability of a chain configuration is a Boltzmann distribution. Each pair of monomer interacts with a spherical-well potential of hardcore radius $r_{\mathrm{HC}} = 0.6a$ and range $R = 1.5a$. The value of R and r_{HC} were determined to give the best agreement between calculated and experimentally observed contact probabilities during the optimization procedure of the potentials B_{ij} (Fig.33C).

The output of the simulations based on the potentials constructed from the coefficients B_{ij} [270] is used to produce polymer configurations compatible with experimental data (DNA FISH)[91]. Cell-to-cell variations in chromosome structure are correlated at a single cell level with the ones in transcription levels of Tsix and other transcripts within the same TAD [91]. These variations predict genetic modifications such as the deletion of beads from the polymer, which are verified experimentally. Finally, interactions within each TAD contribute to the stability of the boundary between them. To conclude, a common effective long-range potential acting on dispersed binding monomer sites along a polymer chain is sufficient to generate micro-segregation, TAD and sharp boundaries [241]. However, this long-range interaction do not have yet a direct physical support and TAD can also result from long-range interactions due to binding proteins position at random position inside TADs.

It might be possible that TADs can be generated by different physical mechanisms. Supercoiling induced during transcription [147] or unconstrained supercoiling in chromatin [191, 146] can be generated in models by taking a rod, twisting it ΔLk times and gluing the two ends together. The resulting loop has to coil in order to relieve tension [36]. TADs could also emergence from supercoiling [23] and in between boundary elements, there are sections of unconstrained chromatin which is super-coiled. In this model, the axial rotation of chromatin fibers is limited, although the anchoring nuclear granules are free to move. Strong supercoiling with $\Delta Lk = -8$ per 400kbp reproduces contact probabilities occurring within individual TADs and between neighboring topological domains in undifferentiated embryonic stem cells [197]. Furthermore, weak supercoiling ($\Delta Lk = -1$ per 400kbp) is used to reproduce the contact probabilities in differentiated cells [23].

6.4 Chromatin loops extrusion as a mechanism for chromosome segregation and TADs

Mitotic chromosomes are condensed by enzymatic proteins (condensin) that actively extrude chromatin loops [190, 6]. Condensin and cohesin complexes play a central role in chromosome compaction [231, 115, 185], but direct experimental evidences are still lacking. The enzymes link two adjacent sites and move both contact points along the chromosome in opposite directions, so that they progressively bridge distant elements stalling at specific sites (CTCF). A polymer model was constructed based on loop-extruding factors (LEFs), where two heads bound to nearby sites along the chromatin fiber can slide away from each other [92], leading to a loop extrusion (fig.36a). When the heads of two neighboring LEFs collide, they block each other and stop. LEF can dissociate with a certain rate independent of their state and location and rebinds immediately at a random site elsewhere on the chromosome, where it resumes extruding a new loop (Fig.36a). Thus, LEFs can generate a tightly stacked loop array.

Numerical simulations shows that the steady-state does not dependent on initial configurations.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

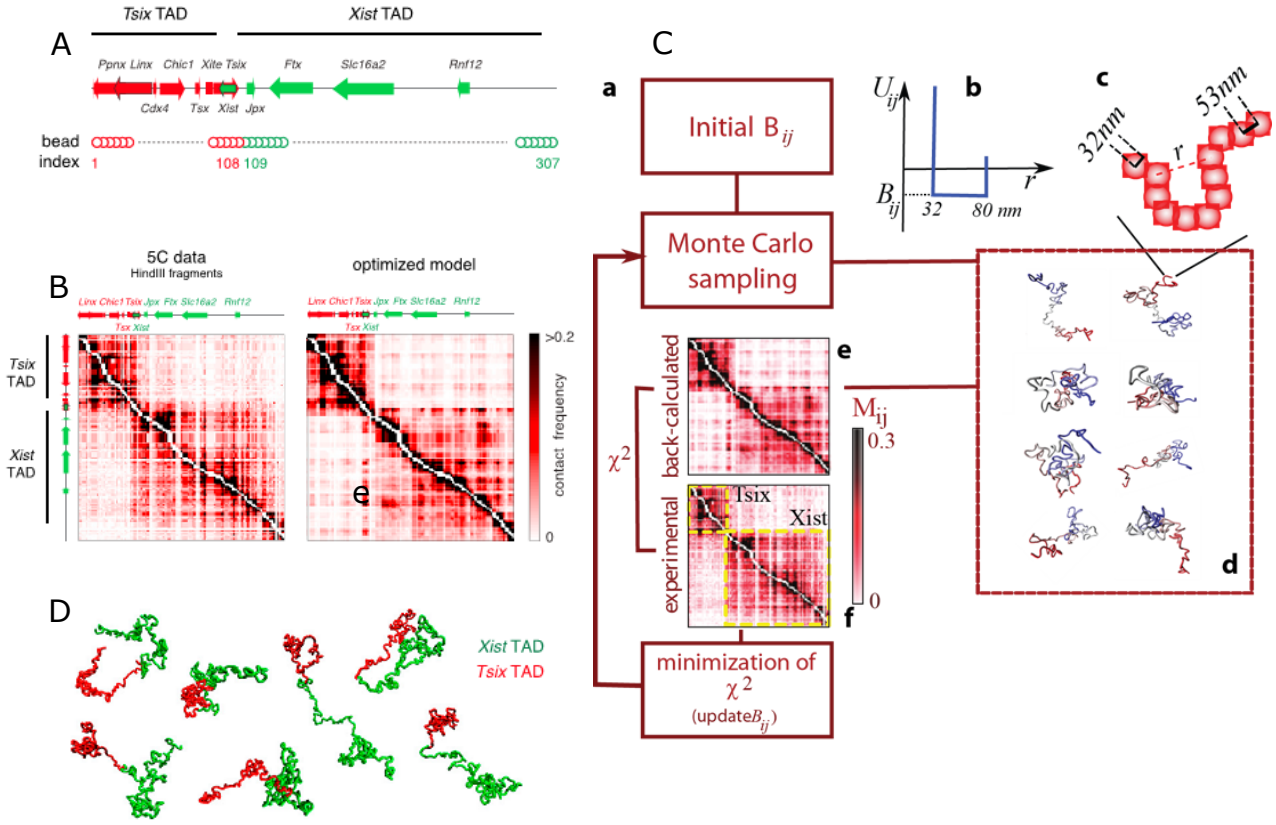


Fig. 33: **Chromosome partitioning into topologically associating domains (TADs).** (**A**) Model of polymer fiber for simulating the Tsix and Xist TADs together (red and green respectively). (**B**) Experimental and simulated contact frequencies for the Tsix and Xist TADs. The model reproduces the two TADs and the weak contact frequencies between them [197]. (**C**): (a) Computational algorithm to determine the potentials that reproduce the experimental 5C contact map. (b) The interaction potential is the sum of two terms shaped as spherical wells of energy B_{ij}. (c) the chromatin fiber is modeled as a chain of beads, each representing a segment of 3 kbp along the chromatin fiber. (d) following each iteration of the algorithm, a Monte Carlo sampling records the conformations computed from the energies B_{ij}. (e) Contact map back-calculated from the sampled conformations. (f) Experimental contact map returned by the 5C experiments, where the Tsix and Xist TADs can be identified and compared with the back-calculated map through evaluation of χ^2 (the squared distance between the experimental and numerically estimated contact probabilities). χ^2 is minimized by updating the energies B_{ij} and re-injection to compute the conformations sampled in (d). Monte Carlo sampling is carried out and the algorithm is iterated until the sequence of χ^2 converges. (g) Simulated vs. experimental contact maps for the Tsix and Xist TADs in mouse ESCs (adapted from [91]). Arrowheads indicate the frequent interactions between Linx, Chic1 and Xite described in [197] (adapted from [270]). (**D**) Sample conformations from the optimized simulation shown in (B), compartmentalization of the model fiber into two separated domains corresponding to the Tsix and Xist TADs despite extensive structural variability (adapted from [91]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Using this scenario, loops, sub-TAD contact domains and TADs are generated through loop-extruding factors (Fig.36). Upon deletion of a domain boundary, topological domains extend beyond their boundary. The encounter probability map of CTCF binding sites contains also isolated peaks [214, 280]. This scenario is in contrast with the looping interaction models, mediated by protein binding used to model TADs formation, as discussed in previous sections see also [84, 92].

Furthermore, numerical simulations of polymer with LEFs reveal two distinct steady-states of loop-extrusion dynamics: at low LEF concentration, polymers are poorly compacted, where loops are formed by single LEFs and separated by gaps. At high concentration, the chromosome is compacted into an array of consecutive loops, having multiple LEFs (Fig.36c). For a LEF abundance of one condensin per $10 - 30$kb, the loop sizes are consistent with the ones inferred from Hi-C data [192]. These results further suggest that chromosome compaction occurs in the dense regime (Fig.36d). Large scale Hi-C data of contact maps of the human genome with a 1kb resolution [214] reveals the existence of 10,000 loops with an average size of 185 kb, occurring between CTCF motifs. The encouter probability P_{enc} peaks at 5kb and thus the Kuhn length is less than 5kb, measured experimentally in [231]. The decay of $P_{\text{enc}} \sim n^{-1.27}$ is computed from data, which is faster than previously suggested [157]. Within a topological domain, the decay is slower ($P_{\text{enc}} \sim n^{-0.75}$), as expected due to higher interactions (see 5.4.2).

Recently a new method was developed to construct a minimal polymer model from the encounter probability (EP) distribution between genomic sites represented in a large matrix [246]. Although this matrix is obtained by averaging the EP over cell population, the TAD diagonal blocks contains hidden information about sub-chromatin organization. To construct the polymer, the EP decay is used in two steps: first to account for TADs, random connectors are positioned inside a restricted region defining the TADs. In the second step, the long-range frequent specific genomic interactions in the polymer architecture are taken into account for deriving from the EP matrix the strength of the interaction. Interestingly, only a small number of randomly placed connectors are required to reproduce the EP of TADs. The mathematical difficulty solved in [246] is to determine the strength and the number of the connectors, so that the EP of the data and the simulations have same decay exponent α in the formula $P_{\text{enc}} \sim n^{\alpha}$ (Fig. 34). The advantage of this model is that connectors can be directly interpreted as direct molecular binding.

Following the construction of a polymer model, stochastic simulations can be used to compute the distribution of first time and the conditional encounter probability of any genomic sites to meet. Using a polymer with 307 monomers, the encounter time and the probability were estimated for monomer 26 (position of the Linx) to meet monomer 87 (Xite) before monomer 64 (Chic1). In this coarse-grained model, each monomer represents three key sites on the X chromosome [91] located inside TAD D. Three polymer realizations indicate the location of the three sites inside TAD D (yellow) and E (blue), that do not mix (Fig. 35). Numerical simulations reveal that the encounter probability is $P = 0.55$, while the mean times are quite comparable of the order of 131s (see table C in Fig. 35). Interestingly this approach can be used to check the impact of TAD E on the mean encounter time inside TAD D: after removing TAD E, stochastic simulations (Fig. 35D) show that the encounter probability was inverted compared to the case of no deletion: the mean encounter has increased by almost 50% to 195s (Linx to Chic1) and 205s (Linx to Xite). This result suggests that TAD E contributes in modulating the interaction probability and the mean time and thus further indicates that the search time inside a TAD depends on neighboring chromatin configuration, due to interaction between TADS present in the data. These encounter times reveal how chromatin can self-regulate. The present polymer construction is generic and can be used to study steady-state and transient properties of chromatin constrained in 5C data.

To conclude, the models described in this section assume that the polymer fiber is at equilib-

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

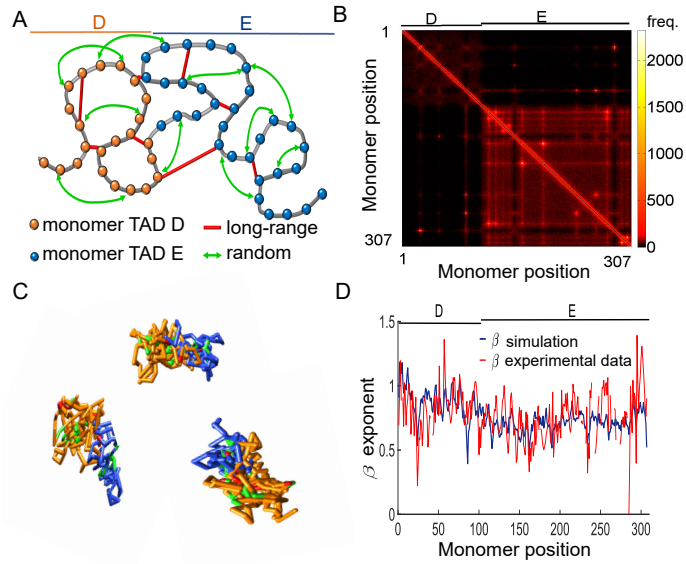
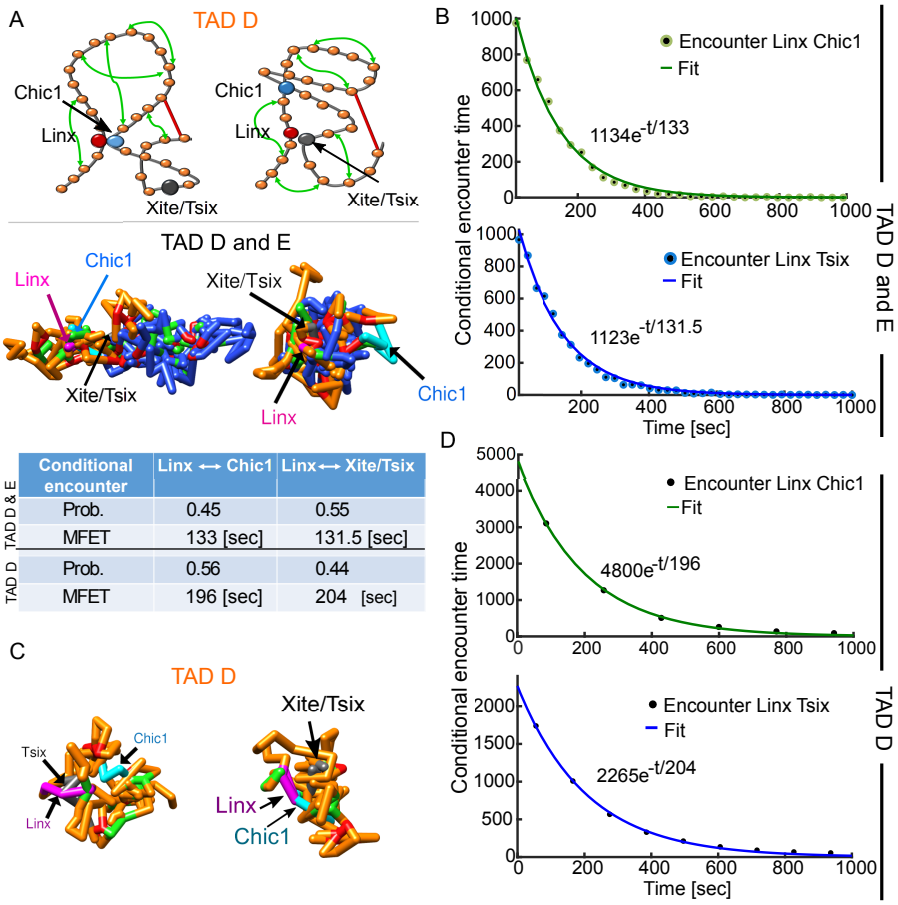


Fig. 34: **Coarse-grained reconstruction of chromatin using extracted random loops and connectors peaks present in 5C data**. **A.** Schematic polymer model where TAD D (orange, monomers 1-106), and TAD E (blue, monomers 107-307) are recovered from random loops (green arrows) according to the connectivity ξ and fixed connectors (red bars) corresponding to the specific peaks of the 5C data. **B.** Three realizations of the polymer model. **C** Encounter frequency matrix of the simulated polymer model, showing two TADs, and the off-diagonal points correspond to the fixed connectors. **D.** Comparison between β computed from experiments and simulations data, confirming that the present polymer model accounts for the statistics of encounter frequency distribution.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



	Conditional encounter	Linx ⟷ Chic1	Linx ⟷ Xite/Tsix
TAD D & E	Prob.	0.45	0.55
	MFET	133 [sec]	131.5 [sec]
TAD D	Prob.	0.56	0.44
	MFET	196 [sec]	204 [sec]

Fig. 35: **Transient properties of the chromatin: Conditional mean time and probability for three sites to meet**. **A.** (upper panel) Representation of the polymer model for TAD D (orange, monomers 1-106), where loci Linx (bead 26, red) meets Chic1 (bead 67, cyan) and Xite/Tsix (bead 87, gray) respectively. Random loops (green arrows) and specific long range-connectors (red bar) are added following the connectivity recovered from data [246]. Fixed connectors (red bars) correspond to the specific peaks of the 5C data. Three realizations (bottom panel) of the polymer model containing TAD D and E, showing the encounter between Linx (magenta) and Chic1 (cyan) and Xite/Tsix (gray) respectively. The color code is from the upper panel. **B.** Histogram of the conditional encounter times between Linx and Chic1 (upper panel, green),and Linx and Xite/Tsix (bottom panel, blue) with TAD D and E. **C.** Two polymer realization with a single TAD D (bead 1-106, orange) showing the encounter between Linx (magenta) and Xite/Tsix (gray, left panel), and the encounter between Linx and Chic1 (cyan, right panel). **D.** Histogram of the conditional encounter times for a polymer with only TAD D, showing an exponential decay as in subfigure C.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

rium. However, one fundamental difference between metazoan and yeast genomes is that while yeast chromosomes can equilibrate during one cell cycle, metazoan chromosomes are so long that they are unable to unmix in the time interval between mitosis [223]. Hence, it is possible that TADs are structures generated during mitosis. The prediction of the models discussed here remain to be tested, especially for the formation and maintenance of TAD boundaries. Coarse-grained polymer models certainly provide a theoretical basis to analyze Hi-C data. As these data become more ubiquitous and the spatial resolution increases, more accurate models would be required to describe this increasing complexity.

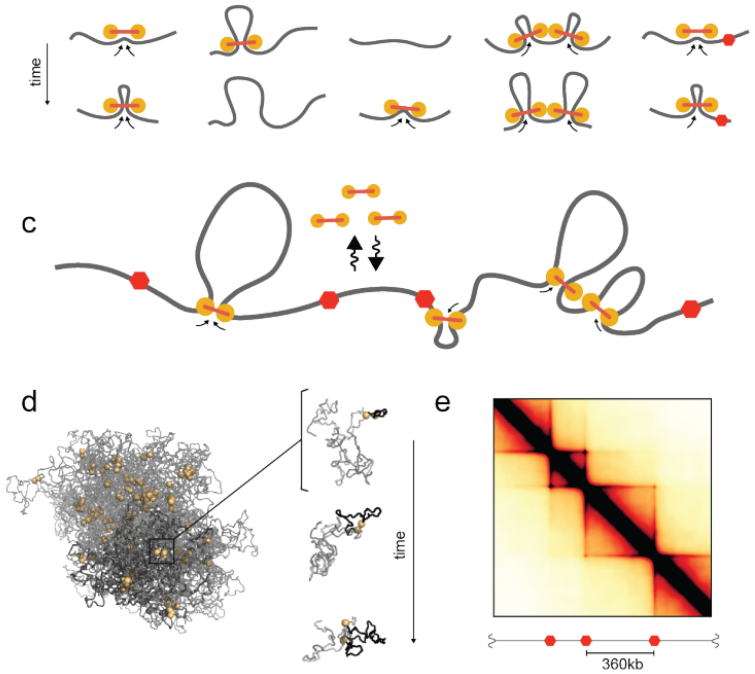


Fig. 36: **Chromosome partitioning into TADs.** (**a**) Model of LEF dynamics, LEFs shown as linked pairs (yellow circles), chromatin fiber (grey). From left to right: extrusion, dissociation, association, stalling upon encountering a neighboring LEF, stalling at a BE (red hexagon). (**b**). Schematic of LEF dynamics. (**c**) Conformations of a polymer subject to LEF dynamics, with processivity 120kb, separation 120kb. Left: shows LEFs (yellow), and chromatin (grey), for one conformation, where darker grey highlights the combined extent of three regions of sizes (180kb, 360kb, 720kb) separated by BEs. Right: progressive extrusion of a loop (black) within a 180kb region. (**d**) Simulated contact map for processivity 120kb, separation 120kb. (adapted from Fudenberg et al. [84]).

7 Coarse-graining polymer models for analyzing telomere organization in yeast nucleus

Telomere organization is a specific characteristic of the nucleus, in addition to the nucleolus, centromeres and different bodies (PML, Cajal) [265]. In budding yeast, the 32 telomeres of a haploid cell can associate in several clusters [94, 263, 225, 124] as shown in Fig. 37. Although a single telomere cannot be continuously monitored, the formation of telomere clusters observed *in vivo* have opened the debate of their role, function and whether these clusters resulted from random encounter without

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

physical interactions or from a dissociation-aggregation process.

In yeast, telomere dynamics lead to the formation of two to six clusters, composed of an average of four telomeres [225]. Telomere foci undergo fusion and fission events over a time-scale of minutes [235]. Numerical simulations underlying telomere dynamics are realized by adding structural constraints, including chromosome structure, attachment to the spindle pole body (SPB), and nuclear crowding [269, 297], suggesting that clusters are due to transient encounter.

Although telomere foci observed *in vivo* could result from random encounters of telomeres at the

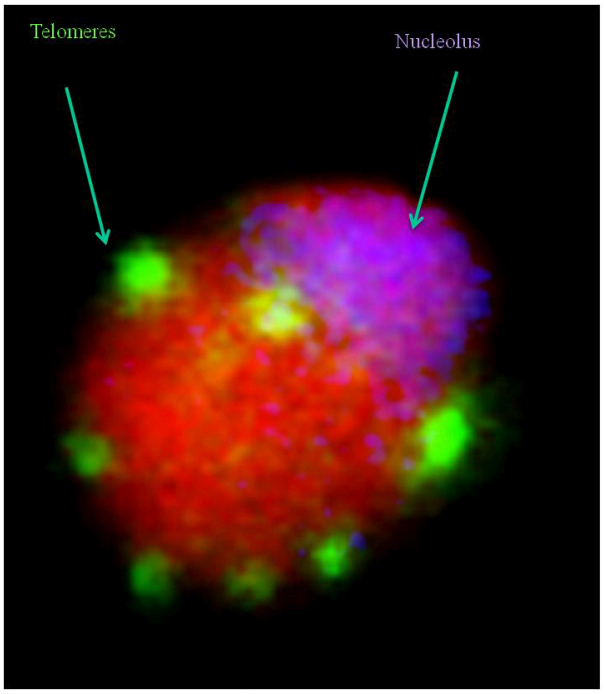


Fig. 37: **64 telomeres of diploid cells.** Telomeres are clustered of 8 to 10 perinuclear foci on which silencing factors concentrate (reproduced from [94]).

nuclear periphery, stochastic simulations of non-interacting telomeres have shown that this hypothesis would contradict the stability of foci observed invivo, where telomere foci are stable over minutes ([235] and Fig. 1G, Movie S1 and S2 of [124]). In numerical simulations, telomere dynamics can be coarse-grained to 32 independent Brownian particles, diffusing on the two-dimensional nuclear envelope of radius 1μm with a diffusion constant $D = 0.005\mu m^2/s$ [39]. The nucleolus region which occupies about 1/3 of the total surface is excluded. This result is in contrast with previous simulations based on microscopy images of fluorescently tagged telomeres, where two telomeres located at a distance $< 0.3\mu m$ were considered to be part of the same focus [297]. Thus, if telomere clusters cannot result simply from the transient encounter of independent telomeres moving by random motion, clustering dynamics and stability are due to aggregation-dissociation processes (see discussion below).

7.1 Encounter rate of telomeres

Coarse-graining the motion of one end of a Rouse polymer (Fig. 38A) into a single Brownian particle moving on the surface of a sphere was recently used to study telomere clustering. Indeed the arrival time of a single telomere moving on the surface of the nucleus to a cluster (which occupies a small fraction of the surface) is a rare event, taking a long time [239]. This coarse-graining approximation

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

was validated by Brownian simulation of a Rouse polymer, which models the chromosome dynamics (Fig. 38). In this simulation, the telomere motion occurs on the surface of a sphere, which represents the nuclear surface, whereas the remainder of the monomers in the chain evolves inside the ball (modeling the nucleus). The distribution of telomere arrival times to a small target, which can be another telomere or a small cluster, is well approximated by a single exponential (Fig. 38B). The encounter time of telomeres at the nuclear periphery is characterized by the arrival time only, even though telomere motion involves complex polymer chains accounting for the physical chromosomal chain. Consequently, telomere clustering is mediated by the arrival of a chromosome to a small cluster and this process is modeled as Poissonian, as long as the chromosome arm does not restrict the motion of the telomere on the nuclear surface.

In summary, the arrival time of a telomere to a cluster can be simulated using a Poissonian distribution approach. To study 32 telomeres (Fig. 38C), it is thus enough to run molecular dynamics simulations of Brownian particles on the surface of a sphere, except on the nucleolus region, where there are reflected. This process is characterized by the reciprocal of the mean encounter time k_f between two telomeres and dissociation process explain later on. The arrival time of a telomere to a cluster is Poissonian (Fig. 38C).

Molecular dynamics simulations of two stochastic particles on the surface of a sphere shows provide a

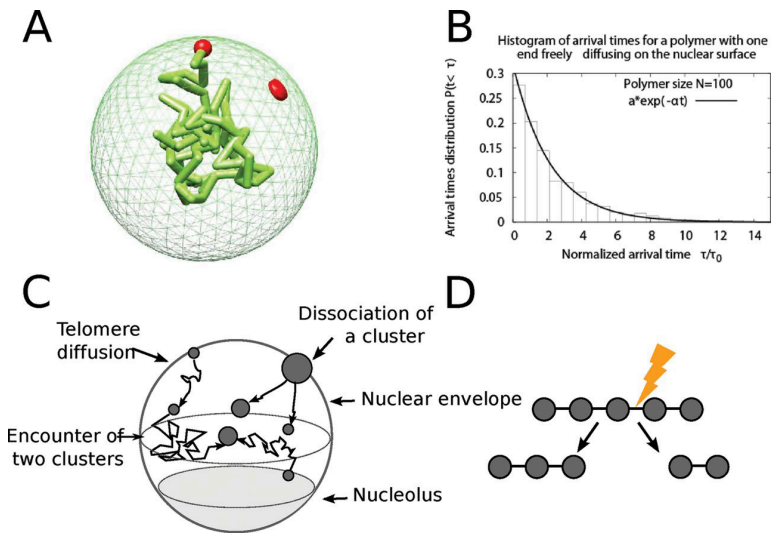


Fig. 38: **Computational model of telomere cluster formation.** (A) Snapshot from a Brownian Dynamics simulation of a polymer with one end anchored on the nuclear surface. The polymer is composed of 100 monomers with average distance between monomers $l_0 = 50$ nm, the nucleus is a reflecting sphere of size $R = 250nm$. (B) Histogram of the arrival times for a polymer of size 100 monomers freely diffusing in the nucleus and one end constrained to the surface. A fit of the form $f(t) = a\exp(-bt)$ gives $a = 0.3$, $b = 0.44s^{-1}$. (C) Scheme of the diffusion-aggregation-dissociation model of telomere organization. Telomeres are represented as Brownian particles diffusing on the nuclear surface, and two telomeres coalesce with a rate k_f and a cluster of n splits at a rate $(n-1)kb$. (D) Schematic representation of the cluster dissociation model, where a cluster of n telomeres has $n-1$ bounds. Any of these bounds can break at a rate k_b and the cluster effective dissociation rate is $(n-1)k_b$ (reproduced from [124]).

mean of estimating the encounter rate of the order $k_f \approx 1.910^{-3}s^1$, for a target which is a disk radius r= 0.015 μm. Thus the encounter rate for two telomeres is k_f. Following an encounter, although

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

telomeres form a cluster that in principle varies in size, because this size remains small, possible change in the scattering cross-section and motility, which could modify the forward binding rates [122] can be neglected. Thus the encounter rate between clusters or telomeres is approximated by a constant k_f, independent of the size.

7.2 Dissociation of a telomere from a cluster

The dissociation of a telomere from a telomere cluster is still not clearly understood, because there are many possibilities for a particle to dissociate from an ensemble. This process has been described using the hypothesis that dissociations are single independent Poissonian events [122, 124], where a cluster containing n telomere dissociate with a rate $(n-1)kb$, where kb is the dissociation rate between two telomeres (Fig. 38D). In addition, dissociation gives rise to two clusters of random size p and n-p, and the probability is uniform. In the absence of specific information about the molecular organization of a cluster, it is difficult to account for the local molecular interactions of telomeres in a cluster. This simple model of dissociation is independent of any polymer model, but could be added to any future polymer models. A dissociation model based on polymer model is still to be formulated, but should involve the SIR family of proteins (A. Taddei, private communication).

The clustering of telomeres was studied using numerical simulations as follow: free telomeres can bind together to form clusters with a forward rate k_f or dissociate with a backward rate k_b. The ratio of these two constants defines an equilibrium, characterized by the ratio parameter $a = k_b/k_f$. When association and dissociation rates are Poissonian, the classical Gillespie's algorithm can be used to simulate telomere dynamics and study the cluster distributions. By comparing experimental distribution [225] with simulations [124], under the dissociation model described above, the dissociation rate of a telomere pair was found to be $k_b = 2.310^{-2}s^{-1}$. Thus, the mean lifetime of a bond between two telomeres is $1/k_b = 43.5$sec. Interestingly, under a genetic perturbation where a fundamental protein to link telomeres together (the silencing factor Sir3) has been overexpressed by 6- or 12-fold the endogenous level, then the backward rate changed to $k_b^{-1} = 109$min for GALSp and $k_b^{-1} = 154$min for GAL1p [124]. The exact molecular changes underlying these differences is still lacking.

7.3 Characterization of telomere exchanged between clusters

We have seen in the previous section that a combination of polymer models with live cell imaging led to the conclusion that telomere organization in wild-type cells results from several physical processes: aggregation mediated by direct interactions between telomeres, dissociation resulting from the separation from a cluster, and telomere random motion located on the nuclear envelope. Telomere Clusters are constantly reshaped due to binding and unbinding, leading on average to three to five detectable foci of four telomeres [94, 225, 122, 124]. Although, cluster dynamics is characterized by the exchange rate of telomeres, two quantities are conserved in this process: the first one is the co-localization time T_C, which is the time that two telomeres spend in a cluster (including cluster of two telomeres), when other telomeres can be exchanged (detached or attached). This time is not equal to the classic dissociation time, which represents the lifetime of a bond between two telomeres, because it accounts for all events of binding and unbinding leaving the two telomeres inside the same cluster. Using Brownian simulations in yeast, the co-localization time was estimated T_C $23.4s$ [124].

The second conserved quantity to characterize clustering is the recurrence time T_R, defined as the mean time for two telomeres to meet again in a cluster after they have just been separated. Two telomeres return in the same cluster in a mean time T_R (T_R is the sum of the time that both telomeres spend separately in different clusters plus the time to travel between clusters until they meet again for

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the fist time) of the order of 480sec, whereas they stay together in a given cluster for 23sec only [124], consistently with experimental observations [269]. The probability P_2 to find two given telomeres in the same cluster (including in a pair) is equal to the ratio of times, $P_2 = T_C/(T_C + T_R)$, where T_C is the mean time that two telomeres spend in the same cluster. This probability is $P_2 = 0.047$, consistent with the result of [269], where the probabilities for two telomeres to belong to the same cluster were determined experimentally to be mostly in the range $0.04 - 0.09$ (Fig. 39).

In summary, coarse-grained polymer models with live cell imaging led to the conclusion that telomere

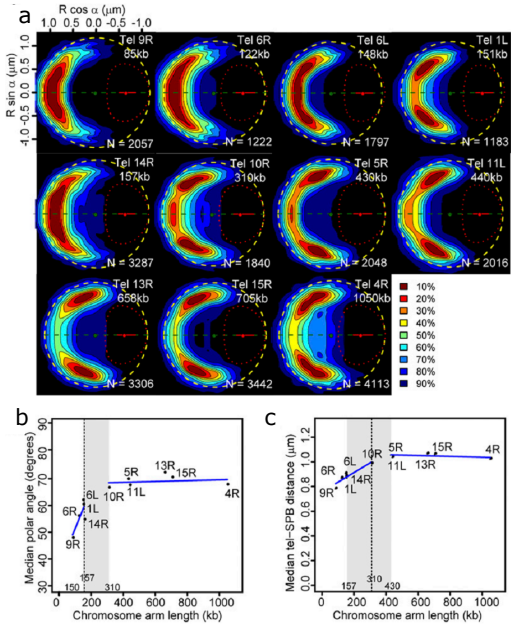


Fig. 39: **Arm length influences localization of subtelomeres at the nuclear periphery.** (**a**) 2D localization probability maps for different telomeres. Top and bottom halves are mirrored (around the green dashed line) for visual convenience. Yellow circle and red dotted outline represent a median nuclear envelope and nucleolus, respectively. The probability to find a locus inside different regions of the nucleus is indicated by the percentage of the enclosing contour. Hot colors indicate higher probability densities. Chromosome arm sizes are indicated on the right (kb). N, number of cells analyzed. (**b**) Median polar angle vs. chromosome arm length, annotated with the subtelomere label. Dashed vertical line represents the change point value, gray area represents its 95 percent CI. Linear regression relationships are $Y = 46.3 + 6.110^{-2}$ L (kb), for arm length $L = 430$kb and $Y = 85.5 + 1.910^{-2}$ L (kb), for L larger than 430 kb. (**c**) Median subtelomere-SPB distances calculated as $\text{dist(Tel, SPB)} = \sqrt{R_{Tel}^2 + R_{SPB}^2 + 2R_{Tel}R_{Tel}\cos a}$, where R_{Tel} is the distance of subtelomere to the nuclear center, R_{SPB} is the distance of SPB from the nuclear center = 900 nm, and a is the angle with the horizontal axis. Dashed vertical and gray areas are defined as in b (reproduced from [269]).

organization in yeast wild-type cells results from three physical processes: aggregation mediated by direct interactions between telomeres, dissociation resulting from the separation from a cluster, and telomere random motion located on the nuclear envelope. Clusters are constantly reshaped due to binding and unbinding, leading on average to three to five detectable foci of four telomeres. An exact model of telomere dissociation in yeast is still missing. Including super-resolution imaging information about the local geometry would be necessary. It would also be interesting to study telomere dynamics in other species.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

8 Single particle trajectories of a chromatin locus

In this section, we describe the statistical analysis of Single Particle Trajectories (SPTs) of a chromatin locus. Many of the basic cell functions such as transcription, repair and division can now be monitored in vivo, but the interpretation of the trajectories remains difficult due to all mechanisms involved in generating the motion. There is also a large unexplained variability. We first describe how SPTs data are acquired and in the second part, we present recent polymer models and their statistics used to interpret data and extract biophysical parameters.

8.1 Live imaging of nuclear elements

The data acquisition for the motion of a chromatin locus is based on using a green fluorescent protein (GFP), extracted from luminescent jellyfish Aequorea victoria. This protein is excited using wavelengths 395nm and 475nm *in vivo*. In that case, processes such as transcription [43] can be monitored in cells. This discovery had a huge impact in cell biology, revealing cell function at a molecular level, as described by The Tsien's 2008 Nobel lecture $http : //www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/tsien - lecture.html$).

A genetic construction designed to visualize a given locus is to integrate a Lac operon at different positions in the yeast genome [220] and using the LacI protein bound to a GFP (making a LacI-GFP complex). This type of molecular construction (Fig. 40a) permits a chromatin locus to become visible (Fig. 40b) and for example sister chromatin separation in yeast can be visualized [180]. When a GFP is fused to nuclear pore proteins (Nup49p) located in the nuclear membrane [111], the motion of the nuclear envelope can be monitored. When both nuclear pores and a locus are labeled, their relative motion can be observed. Once nucleus elements are labeled, high resolution microscopy [243] is used to monitor DNA locus motion at tens of nanometer precisions. Another issue of image acquisition is the possible motion, division of cells or their change to an environment, that should be accounted for in single locus analysis. For example, when a protein diffuse with a diffusion coefficient of the order of $1\mu m^2/sec$ [126] (although there is large variability depending on the protein), the diffusion of a chromatin section can be three orders of magnitude slower [170]. Thus, capturing cellular processes requires a time resolution of seconds or below.

Once nucleus elements are labeled, high resolution microscopy [243] is used to monitor DNA locus motion. Using deconvolution procedure based on the Point Spread Function (bright fixed spot spread of a Dirac) [86, 177], single particle trajectories are acquired, but contain a noise due to the imprecision in the location. Deconvolution procedures further introduce various localization errors that must be differentiated from physical effects [116]. Although tracking procedures remain an image processing challenge, some routine algorithms are now implemented in freely available softwares such as ICY [56], ImageJ [234, 232] or Volocity [151].

8.2 Statistics analysis of SPTs

The ability to follow single locus located at different positions on chromosomes [39] revealed the heterogeneity of the nuclear organization. In the yeast *Saccharomyces Cerevisiae*, chromosomes are non homogeneously distributed, but organized in a Rabl configuration [213] (Fig. 41a), where centromeres are attached together. Similarly, telomeres (ends of the chromosomes) forms clusters on the yeast membrane [265]. In these conditions, the statistics of SPTs of chromatin locus reflect the nuclear heterogeneity. Trajectories can be restricted to a small region (fig.41b-i where the confined ball has a radius of 221nm). In fig.41b-ii, although the recording duration is the same as in fig.41b-i, the trajectory was contained in a smaller ball of radius 164nm. The large heterogeneity of loci

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

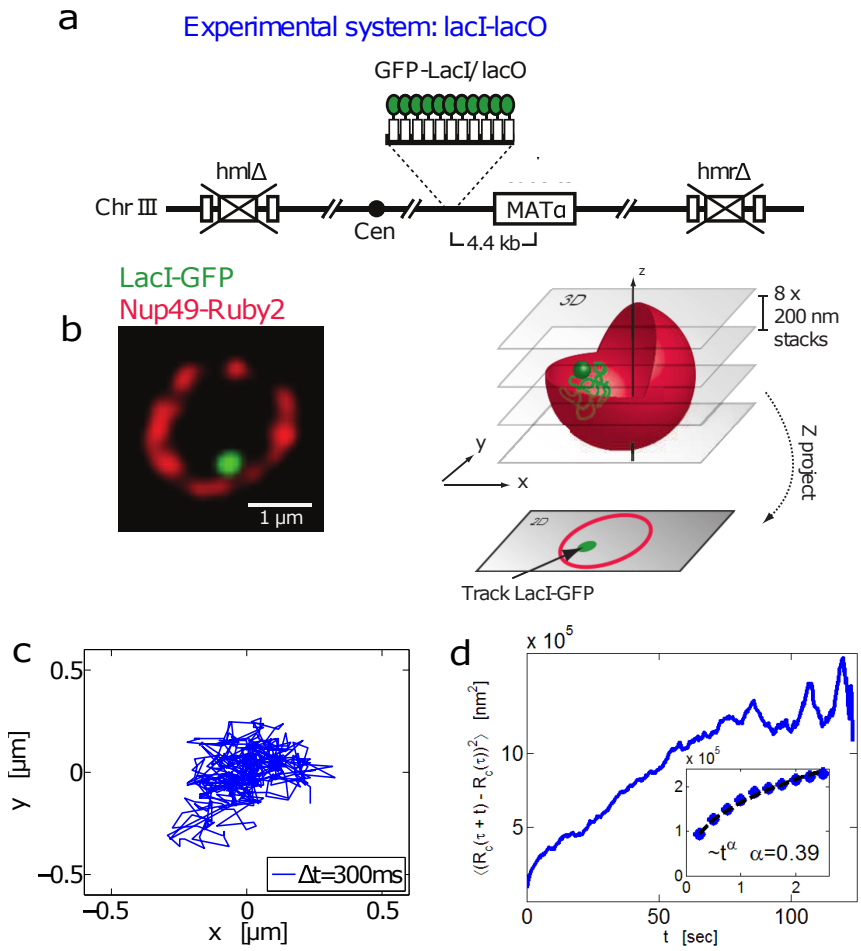


Fig. 40: **Imaging a chromatin locus.** (**a**) LacI-lacO system [188]: Galactose induces a DSB at MAT locus in yeast in a haploid strain that bears lacO sites 4.4 kb from MATa and expresses GFP-lacI and Nup49-GFP fusion proteins (GA-1496). (**b**) MAT locus in yeast (green) while the nuclear membrane (red) was marked with the nup49-mCherry fusion protein (left). The nucleus is imaged in 8 stacks along the z-axis and trajectories are projected on the xy-plane. (**c**) A trajectory of the MAT locus taken at a time resolution of 300msec during 300 frames. (**d**) Correlation function $C(t)$ (eq.198) of the position depicted (from c). At short times, the sub-diffusion regime is characterized by an anomalous exponent $\alpha = 0.39$ (inset magnification at short times) (reproduced from [15]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

behavior across a cell population suggests that the loci can experience different interaction over a time scale of minutes. Analysis of chromosome loci SPTs reveals (Fig. 41b) the local nuclear organization: subtelomeres are positioned at the nuclear periphery, depending on the chromosome arm length, centromeres are attached to the microtubule organizing center (the SPB), and the volume of the nucleolus, which acts as a physical barrier.

To extract biophysical information from SPTs of a chromatin locus, a physical model of the locus motion is needed, based on the local organization of the medium, paved with obstacles and local forces. The classical Langevin's equation 12 and its overdamped approximation 13 usually provides such framework.

The most common statistical estimators to study stochastic trajectories are the Mean Squared Displacement (MSD) and the cross-correlation function, defined for a time series $\boldsymbol{R}_c(t)$ (locus position at time t) by [237, 211]

$$C_\tau(t) = \langle(\boldsymbol{R}_c(\tau+t) - \boldsymbol{R}_c(\tau))^2\rangle, \tag{190}$$

where $\langle . \rangle$ denotes the average over ensemble realization. When the data contains enough recursion, the ergodicity assumption says that the function $C(t)$ can be also computed along a trajectory from the empirical estimator

$$C_\tau^T(q) = \frac{1}{N_p - q}\sum_{h=1}^{N_p-q}(\boldsymbol{R}_c(h\Delta t) - \boldsymbol{R}_c((h+q)\Delta t)^2, \tag{191}$$

for $q = 1, .., N_p - 1$, where $t = q\Delta t$ and $T = N_p\Delta t$ is the duration of the entire trajectory. Fig.40c-d show the function $C(t)$ computed for a chromatin locus. Similarly the MSD of a trajectory is defined by

$$\mathrm{MSD}(t) = \langle(\boldsymbol{R}_c(t) - \boldsymbol{R}_c(0)^2\rangle. \tag{192}$$

The MSD is by definition computed by averaging over $M-$ different trajectories and by taking the limit

$$MSD(t) = \lim_{M\to\infty}\frac{1}{M}\sum_{j=1}^{M}(\boldsymbol{R}_c^j(t) - \boldsymbol{R}_c^j(0))^2, \tag{193}$$

which again assumes that the space is homogeneous. For a Brownian particle moving in $\mathbb{R}^d$ space, characterized by the diffusion coefficient D, the MSD is $\mathrm{MSD}(t) = 2dDt$ [237]. While $C(t)$ is linear in time for a free Brownian particle, it reaches an asymptotic value when the particle is moving in a confined space Ω, with an exponential rate. Indeed, the MSD is expressed in terms of the pdf p_t, solution of the Fokker-Planck equation with reflecting boundary conditions on $\partial\Omega$ and equals to the Dirac-distribution at $t = 0$: $p_0(\boldsymbol{x}, R_c(0)) = \delta(\boldsymbol{x} - R_c(0))$,

$$\mathrm{MSD(t)} = \int_\Omega |\boldsymbol{x} - \mathrm{R_c}(0)|^2 \mathrm{p_t}(\boldsymbol{x}, \mathrm{R_c}(0))\mathrm{d}\boldsymbol{x}. \tag{194}$$

Using an expansion in eigenfunctions $p_t(\boldsymbol{x}, R_c(0)) = \frac{1}{|\Omega|} + \sum_{k\geq 1} e^{-\lambda_k t}\phi_k(\boldsymbol{x})\phi_k(R_c(0))$, where λ_k are the eigenvalue of the Laplacian,

$$\mathrm{MSD}(t) = \frac{\int_\Omega |\boldsymbol{x} - \boldsymbol{R}_c(0)|^2 d\boldsymbol{x}}{|\Omega|} + \sum_{k\geq 1} e^{-\lambda_k t}\int_\Omega |\boldsymbol{x} - \boldsymbol{R}_c(0)|^2\phi_k(\boldsymbol{x})d\boldsymbol{x}.\phi_k(\boldsymbol{x})\phi_k(R_c(0)) \tag{195}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For a disk, $\int_\Omega |\boldsymbol{x} - \boldsymbol{R}_c(0)|^2 d\boldsymbol{x} = \frac{R^2}{2} + |\boldsymbol{R}_c(0)|^2$, and $\lambda_k = D\frac{\alpha_{1k}^2}{R^2}$. The α_{nm} are the ordered m-positive roots of $J_n'(\alpha_{nm}) = 0$, which is the Bessel function of the first kind of order n [26]. The long-time asymptotic for the average MSD(t) over the initial point is approximated by truncating the sum to the first term:

$$< \text{MSD}(t) >_{\boldsymbol{R}_c(0)} = R^2(1 - \frac{8}{\alpha_{11}^2(\alpha_{11}^2 - 1)} e^{-D\frac{\alpha_{11}^2}{R^2}t}). \tag{196}$$

For a general domain, the influence of the domain geometry on the MSD is far more complicated.

For a particle confined in a harmonic potential by a spring force with constant k [237] and friction coefficient γ, the stochastic description is OU-process

$$d\boldsymbol{R}_c = -\frac{k}{\gamma}\boldsymbol{R}_c dt + \sqrt{2D} d\boldsymbol{w}, \tag{197}$$

and the cross-correlation function is

$$C(t) = \frac{2dk_BT}{k}\left(1 - e^{-t/\tau}\right), \tag{198}$$

with $\tau = \gamma/\kappa$, while

$$\text{MSD}(t) = \frac{dk_BT}{k}\left(1 - e^{-2t/\tau}\right), \tag{199}$$

where $\tau = k_BT/D\kappa$. We note that the auto-correlation function for an OU process is

$$\langle(\boldsymbol{R}_c(t) - \langle\boldsymbol{R}_c(t)\rangle).(\boldsymbol{R}_c(t') - \langle\boldsymbol{R}_c(t')\rangle)\rangle = \frac{dk_BT}{k}(e^{-(t-t')/\tau} - e^{-(t+t')/\tau}). \tag{200}$$

In summary, when the correlation function $C(t)$ converges to an asymptotic value, the underlying motion can either be restricted by a tethering force or by obstacles.

8.3 Second moment statistics of SPTs

The MSD was historically computed by averaging the MSD over a cell population and the effective diffusion coefficient is obtained by evaluating the tangent of the MSD at the origin. Due to the large heterogeneity across cells, averaging over a cell population does not reflect diffusion in a single cell. The confinement radius R_C of an abstract disk, where the locus is confined can be extracted from formula 195 as the maximum of the MSD. The two parameters (D and R_C) have been estimated with and without a dsDNA break [69] and also when some proteins have been deleted [170]. For the yeast *HIS3* locus [193], the diffusion coefficient is $\langle D\rangle = 2.1 \pm 0.2 \times 10^{-3}\mu\text{m}^2/\text{sec}$, while the radius of confinement is $\langle R_C\rangle = 0.61 \pm 0.03\mu\text{m}$. After treating the E. Coli DNA with sodium azide, which inhibits ATP synthesis, the diffusion coefficient of a DNA locus is reduced [283]. Thus, these estimators can be used to characterize the dynamics and localization of a chromatin locus. The extraction methods of the two parameters are mostly relevant to analyze particles described as Brownian, but is it insufficient for a DNA locus moving a confined environment? we shall explore these questions in the next sections.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

8.3.1 Constraint length of a locus

A locus localization can be estimated using the standard deviation (SD) of the position with respect to its mean averaged over time. This SD is a characteristic length, called the constraint length L_C, estimated by the empirical sum

$$L_C = \sqrt{\mathrm{Var(R_c)}} = \sqrt{\frac{1}{T}\sum_{h=1}^{T}(\boldsymbol{R}_c(h\Delta t) - \langle \boldsymbol{R}_c\rangle)^2}, \tag{201}$$

where $\langle \boldsymbol{R}_c\rangle$ is the average position of the locus along the trajectory. For a MAT locus in yeast, where trajectories are recorded for hundreds of seconds, the distribution L_C is shown in fig.41c: the histogram of L_C shows a large variability across cells with an average $\langle L_C\rangle = 0.202 \pm 0.044\mu$m, when the yeasts grew in glucose and higher in galactose.

8.3.2 Effective diffusion coefficient of a chromatin locus

The effective diffusion coefficient of a locus can be computed directly from the time series [237] using the estimator

$$D_c \approx \frac{1}{6N\Delta t}\sum_{h=1}^{N-1}(\boldsymbol{R}_c(h\Delta t) - \boldsymbol{R}_c((h+1)\Delta t))^2, \tag{202}$$

where N is the number of time frames. For short time Δt, the locus motion is described as Brownian (subsection 3.10.2) and the diffusion coefficient is computed from eq.202 [11]. Reducing the time step Δt provides a better estimate of the local diffusion coefficient D_c. Applying formula 202 on MAT locus trajectories leads to a distribution of diffusion coefficients (Fig. 41d) with a mean over cell population $\langle D_c\rangle = 9.1 \pm 2.8 \times 10^{-3}\mu\mathrm{m}^2/\mathrm{sec}$, when the cells grow in glucose [18]. When the motion is not driven by diffusion only, the estimator 202 should be replaced by the correlation function (relation 191) and other parameters can be extracted by fitting a curve At^{α}. The interpretation of A remains unclear, while α characterizes the anomalous behavior (see section 3.10.2).

8.4 Anomalous diffusion of a chromatin locus

The motion of a tagged monomer, part of a polymer, is described as subdiffusion for an intermediate time scale (section 3.10.1). As we discusses above, this description applied to a tagged locus located in a chromosome chain. The correlation function (eq.191) of a DNA locus, fig. 40d) shows the anomalous diffusion behavior, characterized by an exponent $\alpha < 1$, [178, 139, 284].

The anomalous exponent of a locus contains physical information about the underlying polymer model describing the chromatin. However, there is still a debate about the precise meaning of the anomalous exponent computed from empirical data. The anomalous exponent α of many loci along chromosome XII show (Fig. 42a) a mean value around 0.5 (Fig. 42b). Consequently, in [105] it was suggested that the Rouse flexible chain model is sufficient to describe chromatin dynamics in yeast. However, a cell-to-cell analysis reveals that there is a large distribution of the anomalous exponent, suggesting that the chromatin structure changes over time and across cell populations [105, 4, 18].

8.4.1 Variability of the anomalous exponent over yeast cell population

The anomalous exponent is estimated from the projection of SPTs on the axis of an orthogonal frame, but which axis should be chosen? Although the anomalous exponent vary significantly across

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

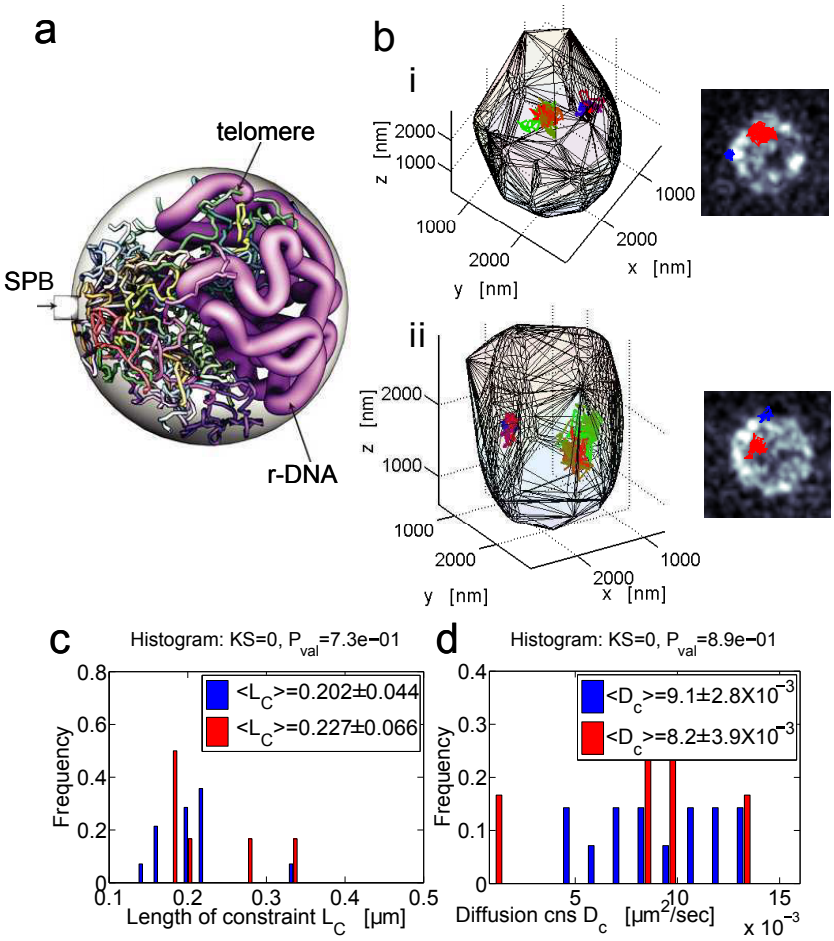


Fig. 41: **Organization of chromosomes.** (**a**) Equilibrium distribution of yeast (Saccharomyces Cerevisiae) chromosome, from Monte-Carlo simulations of polymer [291]. (**b**) Recorded for 100 seconds of two trajectories of a chromatin locus (upper figures-i,ii) and (bottom figures-iii,iv). Time resolution $\Delta t = 0.33s$. At the beginning the trajectory is red and gradually becomes green (locus), while the SPB starts red and gradually becomes blue. The trajectory of the locus (red) and the displacement of the SPB trajectory (blue) inside the nucleus is projected on a plane. The nuclear membrane is reconstructed based on the nup49-mCherry fusion protein. The position and dynamics of the MAT locus in yeast nucleus: (**c**) Distribution of the constrain length (L_C) estimated from eq.201 (100 sec) trajectories when the cells grow in glucose (blue) or galactose (red). (**d**) Distribution of the diffusion coefficients for different cells (eq.202) (reproduced from [18]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cell population, its value is often reported above 0.5 (Fig. 42b,c for yeast). To compare the DNA locus motion with respect to the nucleus, a reference point should be chosen, such as the SPB. Since the SPB is supposed to be fixed at the periphery, any precession (rotational) motion can be detected (Fig. 43a).

The effect of precession on a locus motion is evaluated using Brownian simulations of a polymer confined in a rotating sphere. A polymer end is attached to the center of the sphere. When the angular velocity is $\omega = 0.04\text{sec}^{-1}$ (Fig. 43b), the anomalous exponent increases from 0.5 (Rouse polymer) to 0.75 [18] (Fig. 43c). This result provides a criteria for selecting trajectories originating from different cells: if the SPB is not moving, the anomalous exponent is lower compared to cases where the nucleus is driven by an external motion (Fig. 42c, $\alpha = 0.43 \pm 0.09$). To conclude, the Rouse model characterized by $\alpha = 0.5$ cannot account for the large distribution of anomalous exponent α observed across cell population, which reveals the heterogeneity of the chromatin organization at a given locus site. For $\alpha \leq 0.5$, the $\beta-$polymer model can be used to interpret SPTs and suggests different chromatin organization for a given locus across cell population (see also [246]). Possibly, nucleosomes local remodeling can change the local short- and long-range forces reflected in the distribution of the anomalous exponents [18].

Directed motion has also been reported to affect the anomalous exponent [50] by following telomere motion in ALT (alternative telomere maintenance) cancer cells. In these cells, telomeres elongate via homologous recombination, where they encounter and use each other as a template for elongation: Telomeres perform long-range movement to actively search for recombination partners. Telomeres perform anomalous diffusion with $\alpha = 0.8$ [50] on average. Similarly, the motion of the TetO locus of Tsix TAD in the X chromosome perform intrinsic motion characterized by an exponent $\alpha = 0.56$ [270]. Similar subdiffusive motion are found in pro-B cells [162].

8.4.2 Increasing the anomalous exponent by applying an external drift

Actin network in the cytoplasm can pull the nucleus membrane leading to random oscillations that could potentially affect the internal motion and organization of the chromatin. To explore how a deterministic force acting on a polymer in confined domain affects a monomer motion, numerical simulations have been used where a deterministic force is added to the polymer model. Starting with a Rouse polymer and adding a constant drift in a given direction of amplitude $|\boldsymbol{v}| = 0.2D/b$, the monomer motions are now described by

$$\frac{d\boldsymbol{R_{i,n}}}{dt} = -D\nabla_{\boldsymbol{R_{i,n}}} U(\boldsymbol{R}) + \sqrt{2D}\frac{d\boldsymbol{w_{i,n}}}{dt}, \tag{203}$$

where

$$U(\boldsymbol{R}) = U_{\text{Rouse}}(\boldsymbol{R}_1, ..\boldsymbol{R}_N) + U_{\text{drift}}(\boldsymbol{R}_1, ..\boldsymbol{R}_N), \tag{204}$$

where U_{Rouse} is defined by eq. 5 and

$$U_{\text{drift}}(\boldsymbol{R}_1, ..., \boldsymbol{R}_N) = -\sum_n \frac{\boldsymbol{v} \cdot \boldsymbol{R}_n}{D}. \tag{205}$$

There are no boundaries in the simulations. How this additional motion affects the statistics of a single locus? this question is addressed by computing the anomalous exponent α in the intermediate time regime (Fig. 43d). Contrary to the classical Rouse model or β-polymer models, under a constant pulling force, the anomalous exponent increases to $\alpha = 0.66$ (Fig. 43d). In the long-time regime, the

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

motion of the center of mass becomes ballistic. Thus, adding a deterministic drift on the motion of a polymer results in an increase of the anomalous exponent [18, 15]. Deriving an analytical formula would certainly help to clarify how the anomalous exponent α depends on an external force. There are almost no analytical results in that direction.

8.4.3 Oscillating forces acting on a monomer modify the anomalous exponent

Adding an oscillating force, that can represent oscillation of the nuclear membrane, affects monomer motion and eventually the anomalous exponent. This effect can be studied by adding an oscillatory potential to a Rouse polymer so that the total energy in eq. 203 becomes

$$U(\boldsymbol{R}_1, .., \boldsymbol{R}_N) = U_{\text{Rouse}}(\boldsymbol{R}_1, .., \boldsymbol{R}_N) + U_{\text{osc}}(\boldsymbol{R}_1, .., \boldsymbol{R}_N), \tag{206}$$

and

$$U_{\text{osc}}(\boldsymbol{R}_1, ..., \boldsymbol{R}_N) = \sum_n (\boldsymbol{A} \cdot \boldsymbol{R}_n) \sin(\omega t + \theta_n), \tag{207}$$

where A, ω are constants and compared to eq. 205, the oscillating velocity is written as $\boldsymbol{v} = \boldsymbol{A} \sin(\omega t + \theta_n)$, with phases θ_n. There are two extreme cases first when all $\theta_n = 0$ for $n = 1..N$ or second $\theta_n \sim \mathcal{U}(0, 2\pi)$ are random variables chosen uniformly distributed. By increasing the amplitude A or the frequency ω the anomalous exponent increases (Fig. 43e). Numerical simulations shows that for $A = k_T/b = 1$ and $W = b^2/D$, where θ_n are randomly chosen, the anomalous exponent increases from $\alpha = 0.5$ to 0.8. [18, 15]. In summary, adding a deterministic motion, which may be an artifact of the measuring system, increases the anomalous exponent α. This reason may explain the very high values of α estimated for loci within of chromosome XII [4] of yeast.

Similarly, the rDNA locus is an active chromatin area and is transcribed by RNA Polymerase (Pol I) and Pol III [204], where about 600 protein coding genes are transcribed Such a domain is expected to be under physical stress, where chromatin is repeatedly re-opened. This dynamics suggests that some external forces might be applied on the nucleus. In [90], an alternative approach to model the effect of ATP-dependent active fluctuations was taken, by considering exponentially correlated colored noise acting on the monomers of a semiflexible polymer. In this case the colored noise also resulted in anomalous motion with an exponent much higher than the one obtained with a Gaussian noise ($\alpha = 3/4$ for a semiflexible polymer). If this activity is modeled by an oscillatory motion (Eq.207), high anomalous exponent ($\alpha \sim 0.8$) are expected. While values higher than 0.5 can be explained by a deterministic motion and forces acting on the chromatin, values smaller than 0.5 reflect chromatin condensation, modeled by the β-polymer model (section 3.5) or viscoelastic nuclear medium (section 3.10.3).

8.5 Probing chromosomal structure and dynamics in bacteria using SPTs

Chromosomal structure and dynamics in bacteria can be explored using fluorescently labeled chromosomal loci in E. coli at a timescale of 0.1100 s [129, 128]. A large heterogeneity in trajectories has been reported containing diffusing and ballistic epochs, with abnormally and elongated regions characterized by superdiffusive motion (Fig. 44). These directed movements were analyzed by computing the anomalous exponent α and the diffusion coefficient along single trajectories. The statistics reveals a large spectrum of anomalous exponent, centered around $\alpha = 0.4$, indicating a higher order connectivity compared to Rouse (see subsection 3.10.2). In contrast with sub-diffusion, the ballistic motions reflect active mechanisms or directed forces, and support the heterogeneous chromosome

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

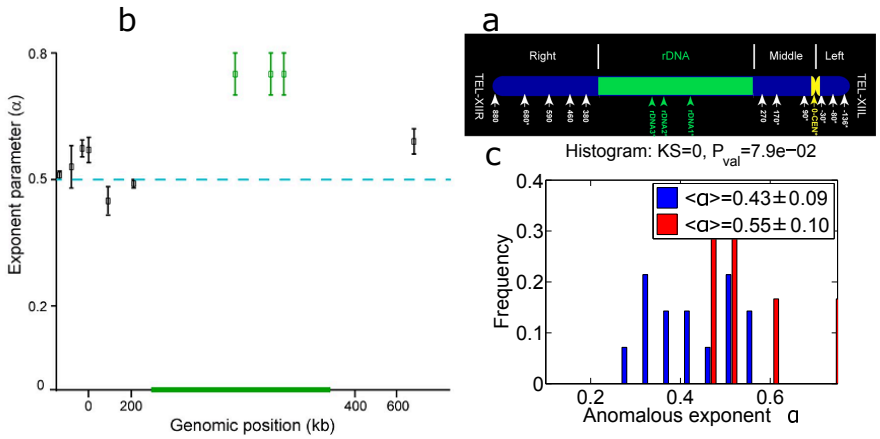


Fig. 42: **Anomalous exponent of chromatin loci in yeast. (a)** Color-coded statistical mapping of positions of 15 loci along chromosome XII. Schematic representation of chromosome XII with the 15 FROS-labeled loci. The rDNA of 1.8 Mb is depicted as a 1-Mb segment. Loci studied by particle tracking are marked with asterisks [4]. **(b)** Anomalous parameter versus the genomic position for 10 loci [4]. **(c)** Distribution of the anomalous exponent from the MAT locus (chromosome III) for chosen cells with a fixed SPB when the cells grow in glucose (blue) or galactose (red) (reproduced from [18]).

motion on a viscoelastic background for the bacterial nucleoid or local forces. To conclude, the variability in the anomalous exponent α reveals the local effect of nuclear proteins on the chromosome motion.

Interestingly, there is also a relation between dynamics and geometrical position: Indeed the loci with the highest variability belong to a macrodomain found near the cell periphery [129], possibly subject to tethering interactions at least during certain parts of the cell cycle. Direct computation using formula 226 should clarify the nature of the ballistic motion.

8.6 Local interactions acting on chromatin

External and internal forces are constantly remodeling the chromatin driven by protein complexes that can compactify and reorganize the DNA. Chromatin fragments interact with many partners such as the nuclear membrane, other chromosomes or nuclear bodies, but the resulting forces cannot be directly measured *in vivo*. However, these forces impact chromatin dynamics and should be reflected in particular in the motion of a single locus (Fig. 45a). It is now possible to extract tethering forces applied on chromatin from the statistics of a single locus trajectories imaged. Indeed, DNA locus plays the role of a probe to reveal the local forces (Fig. 39c). The motion of a chromatin locus can be driven by local diffusion and/or forces between monomers of the model polymer [138, 291, 274]. Monomers motion is highly correlated due the polymer hierarchy of relaxation times [72, 254], leading in particular to anomalous diffusion [137]. Much of the chromatin dynamics is reflected in the motion of a single chromosomal locus and conversely, a locus motion allows probing the chromatin dynamics [154, 177] at tens of nanometers and millisecond scales resolution [297]. Measuring these forces from a single locus trajectories can reveal genomic organization, information about chromosome arm interaction and obstacle organization [279, 15].

When a local interaction is applied on a given locus, it affects the polymer chain, and conversely any monomer motion will be affected by possible interactions exerted along the chain (Fig. 45a).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

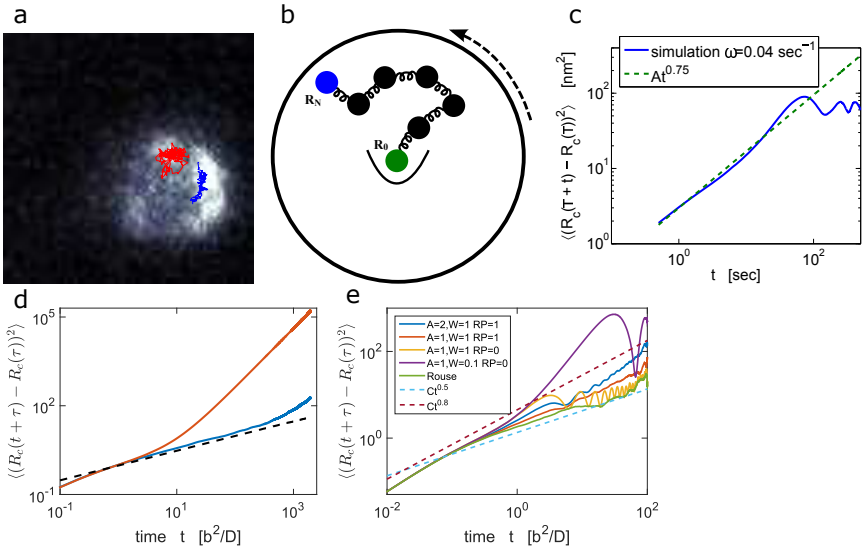


Fig. 43: **Effect of external drift on the anomalous exponent α. (a)** The nucleus is monitored and the SPB (blue trajectory) is performing a precession motion, compared to a tagged locus trajectory (red). **(b)** a Rouse polymer model in a spherical domain, with one end anchored at the center, while following the other end. (**c**) Cross-correlation function and anomalous exponent of the tagged monomer in (b), with $N = 33$ and the domain is rotating in an angular velocity of $\omega = 0.04\text{sec}^1$. The anomalous exponent is approximately equal to 0.75 (and not 0.5 for Rouse). **(d)** Effect of an external drift on the anomalous exponent α: the cross-correlation function was estimated from Brownian simulations of a Rouse polymer ($\beta = 2$ blue line), where the position of the first monomer was followed. At intermediate time, the particle performs anomalous diffusion with $\alpha = 0.66$, which should be compared to $\alpha = 0.5$ (black dashed line). **(e)** The time cross correlation function of the middle monomer position of a Rouse polymer ($N = 33$), where an oscillating force is acting on all monomers (Eq. 207). The phase of forces can be the same for all monomers (RP=0) or random (RP=1). We estimated the cross-correlation function for several values of the amplitude (A) and frequencies (W). Also shown is the result for the Rouse polymer, with no oscillating force, for which the anomalous exponent is 0.5 (green). Two trend lines ($\alpha = 0.5$ and $\alpha = 0.8$) (reproduced from [15]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

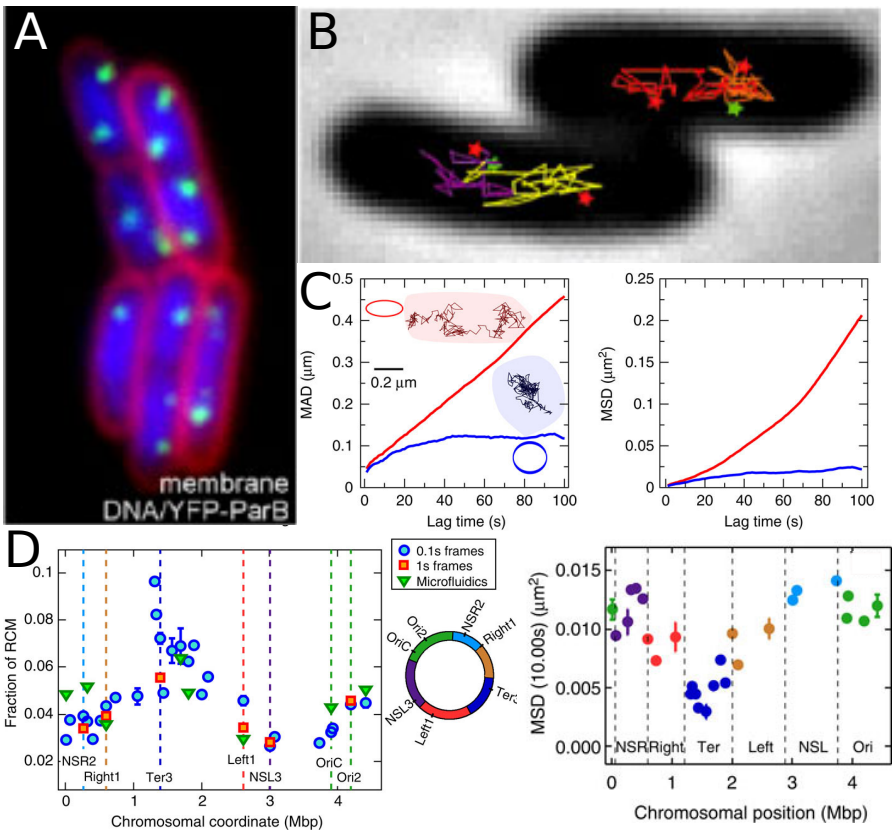


Fig. 44: **SPTs to probe chromosomal structure and dynamics in bacteria.** **(A)** Cell sub-compartment visualization, membrane (FM 4-64), DNA (DAPI)and foci dynamic measurement (YFP-ParB) stainings. (B) Trajectories of fluorescent dots of foci dynamics compared to territories (stars) in which genetic loci are labeled [175]. **(C)** Analysis of chromosomal loci trajectories: a ballistic (red) and a typical (blue) track (Left) are represented with their elliptic convex-hulls. The time-averaged MAD (left) and mean square displacement (right) of the two tracks reveals heterogeneous motions. (D) Fraction of rapid movements (RCM) as a function of chromosomal coordinate: Squares correspond to the loci belonging to different macrodomains. Experiments on 27 loci with frame separation of 0.1s, 1s and in a microfluidics device (triangles). Right: Means square Displacement (MSD) showing the intrinsic differences with the chromosomal location (reproduced from [129] and [128]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

When a far away monomer is anchored, the motion of the observed monomer is certainly restricted. Indeed, when applied forces are stationary over the time course of a trajectory recording, the empirical velocity distribution of a locus is related to forces applied to a distant single monomer. It is possible to distinguish external forces applied on a single monomer from intrinsic forces acting on monomers. The principle and the difficulty of the method can be understood as follows: for a single stochastic particle modeled by the Smoluchowski's limit of the Langevin equation, the velocity of the particle $\mathbf{v}$ is proportional to a force $\mathbf{f}$ applied on the particle plus an additional white noise, summarized as

$$\gamma \mathbf{v} = \mathbf{f} + \gamma\sqrt{2D}\dot{\mathbf{w}}, \tag{208}$$

where γ is the friction coefficient, D the diffusion coefficient and w is the normalized wiener process. Thus by averaging over the ensemble of velocity realizations, it is possible to recover the first moment, which is the force field. However, for a polymer chain, there are internal forces between monomers and thus, as the data are measured at a single monomer, the internal forces acting on the measured monomer need to be separate from the external ones acting on a monomer further away. Analytical formula are constructed to recover the force acting on the polymer model. When the external applied force is the gradient of a quadratic potential. The method is presented in in the next sections, we review recent effort to recover far away interactions from the analysis of SPTs at a given locus.

8.6.1 Recovering a tethering force (external potential) for a Rouse polymer

To extract a tethering force acting on a Rouse polymer from single locus trajectories, we recall first how a force $F(X)$ can be recovered from the trajectories of a single particle, modeled by the overdamped Langevin equation

$$\dot{X} = \frac{F(X)}{\gamma} + \sqrt{2D}\,\dot{W}, \tag{209}$$

where W is a Gaussian white noise and γ is the viscosity [238]. The source of the driving noise is the thermal agitation [238, 236]. The present model and analysis assumes that the local environment where the particle evolves is stationary. Moreover, the statistical properties of the trajectories do not change over time. In practice, the assumption of stationary is usually satisfied as trajectories are acquired for short time, where few statistical changes are expected to occur. The parameters in the coarse-grained model 209 are recovered from computing the conditional moments of the trajectory increments $\Delta X = X(t+\Delta t) - X(t)$ (see [236]),

$$\frac{F(X)}{\gamma} = \lim_{\Delta t \to 0} \frac{\mathbb{E}[\Delta X(t) \,|\, X(t) = x]}{\Delta t}, \quad D = \lim_{\Delta t \to 0} \frac{\mathbb{E}[\Delta X(t)^T \Delta X(t) \,|\, X(t) = x]}{2\Delta t}. \tag{210}$$

Here the notation $\mathbb{E}[\cdot \,|\, X(t) = x]$ means averaging over all trajectories that are at point x at time t.

To generalize relation 210 to a Rouse polymer embedded in an external force, we define the potential $U_{\text{ext}}(\boldsymbol{R})$ and described the total energy of the interacting polymer by

$$\phi(\boldsymbol{R}) = \frac{\kappa}{2}\sum_{j=2}^{N} (\boldsymbol{R}_j - \boldsymbol{R}_{j-1})^2 + U_{\text{ext}}(\boldsymbol{R}). \tag{211}$$

The monomer dynamics follows stochastic eq.14. When monomer $\boldsymbol{R}_c$ is monitored and trajectories are collected, the first moment of the monomer increment position $\boldsymbol{R}_c(t+\Delta t) - \boldsymbol{R}_c(t)$ is proportional to

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the velocity. Computing the increment involves averaging over all polymer realizations and accounts for the entire polymer configuration [17]. The following relation was derived in [17]

$$\lim_{\Delta t \to 0} \mathbb{E}\{\frac{\boldsymbol{R}_c(t+\Delta t) - \boldsymbol{R}_c(t)}{\Delta t} | \boldsymbol{R}_c = \boldsymbol{x}\} = -D \int_\Omega d\boldsymbol{R}_1 .. \int_\Omega d\boldsymbol{R}_N (\nabla_{\boldsymbol{R}_c} \phi) P(\boldsymbol{R}|\boldsymbol{R}_c = \boldsymbol{x}), \tag{212}$$

where $\mathbb{E}\{.|\boldsymbol{R}_c = \boldsymbol{x}\}$ denotes averaging over all polymer configuration under the condition that the tagged monomer is at position $\boldsymbol{R}_c = \boldsymbol{x}$. Formula 212 is generic and generalizes relation 209. Furthermore, it does not depend on the particular expression of the external forces acting on the polymer. No restrictions are imposed on the domain Ω where the polymer evolves, but the polymer is reflected on the boundary $\partial\Omega$. The conditional probability $P(\boldsymbol{R}|\boldsymbol{R}_c = \boldsymbol{x})$ is computed from the equilibrium probability distribution function (pdf) $P(\boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_N)$, which satisfies the Fokker-Planck equation (FPE) in the phase space $\Omega \times ..\Omega \subset \mathbb{R}^{3N}$,

$$0 = \Delta P(\boldsymbol{R}) + \nabla \cdot (\nabla \phi \, P(\boldsymbol{R})), \tag{213}$$

with boundary condition

$$\phi \frac{\partial P}{\partial \boldsymbol{n}_i} + P \frac{\partial \phi}{\partial \boldsymbol{n}_i} = 0 \text{ for } \boldsymbol{R}_i \in \partial\Omega \quad \text{for } i = 1..N,$$

where $\boldsymbol{n}_i$ is the normal vector to the boundary $\partial\Omega$ at position $\boldsymbol{R}_i$. The model for the external force acting on monomer n and located at position $\boldsymbol{R}_n$ is the gradient of a harmonic potential

$$U_{\text{ext}}(\boldsymbol{R}_n) = \frac{1}{2} k (\boldsymbol{\mu} - \boldsymbol{R}_n)^2, \tag{214}$$

where k is the constant force and we choose $n < c$. This external potential well affects the dynamics of the entire polymer and specifically the observed locus c.

8.6.2 Extracting a deterministic force from an ensemble observed locus trajectories

To extract the strength of the potential well applied on monomer n from the measured velocity of locus c, we need an explicit expression for formula 212. First, the force acting on monomer c, when its position is $\boldsymbol{x}$ is given by

$$\boldsymbol{F}^c_{\boldsymbol{R}_c = \boldsymbol{x}} \quad = \quad -\nabla_{\boldsymbol{R}_c} \phi(\boldsymbol{R}_{c-1}, \boldsymbol{R}_c, \boldsymbol{R}_{c+1})_{\boldsymbol{R}_c = \boldsymbol{x}} = -\kappa \left(2\boldsymbol{x} - \boldsymbol{R}_{c-1} - \boldsymbol{R}_{c+1}\right), \tag{215}$$

where ϕ is defined in 211. When the potential well is localized at the origin ($\boldsymbol{\mu} = 0$) in eq.214, the equilibrium pdf is the Boltzmann distribution conditioned on $\boldsymbol{R}_c = \boldsymbol{x}$. It is given by

$$P(\boldsymbol{R}|\boldsymbol{R}_c = \boldsymbol{x}) = \mathcal{N} e^{-\phi(\boldsymbol{R}_1, ..., \boldsymbol{R}_{c-1}, \boldsymbol{x}, \boldsymbol{R}_{c+1}, .., \boldsymbol{R}_N)}, \tag{216}$$

where $\mathcal{N}$ is a normalization factor ($\int_{\boldsymbol{R}|\boldsymbol{R}_c = \boldsymbol{x}} P(\boldsymbol{R}|\boldsymbol{R}_c = \boldsymbol{x}) d\boldsymbol{R} = 1$), involving Gaussian integrals. The result is

$$\frac{1}{\mathcal{N}} = \left[\frac{(2\pi)^{N-1} \kappa^{2-N}}{(\kappa + |c-n|k)}\right]^{3/2} e^{-\frac{\boldsymbol{x}^2(\kappa^2 + (c-n+1)k\kappa)}{2(\kappa + (c-n)k)}}. \tag{217}$$

Substituting eqs. 215 and the normalization expression for $\mathcal{N}$ into eq.212 [17], the mean velocity of monomer c is computed and given by

$$\lim_{\Delta t \to 0} \mathbb{E}\{\frac{\boldsymbol{R}_c(t+\Delta t) - \boldsymbol{R}_c(t)}{\Delta t} | \boldsymbol{R}_c(t) = \boldsymbol{x}\} = \quad -D k_{cn} \boldsymbol{x},$$

$$k_{cn} = \frac{k\kappa}{\kappa + (c-n)k}. \tag{218}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Expression 218 links the average velocity of the observed monomer c to the force applied on monomer n. The coefficient k_{cn} depends on the harmonic well strength k, the inter-monomer spring constant κ and it is inversely proportional to the distance $|n-c|$ between monomers n and c along the chain. Furthermore, the steady state variance $\boldsymbol{R}_c = \lim_{t\to\infty} \boldsymbol{R}_c(t)$ of the monomer's position is given by

$$\langle \boldsymbol{R}_c^2 \rangle = \frac{d}{k_{cn}}, \tag{219}$$

when $\langle \boldsymbol{R}_c \rangle = 0$. Relation 219 is reminiscent of long time asymptotic of classical Ornstein-Uhlenbeck processes. The dynamics of monomer $\boldsymbol{R}_c$ generated by Brownian simulations is shown in fig.45b-c. In the limit of large k (pinned monomer), an analogue of formula 219 was used for analyzing chromatin organization [279] and DNA [158]. Inversion formula 218 assumes the Boltzmann distribution for the single monomer and that the entire polymer has reached equilibrium at the time scale of the simulation or the experiment (from Eq. 216). Finally, formula 218 reveals how internal and external polymer forces are mixing together and are influencing the monomer velocity. It also shows the explicit decay of the force amplitude with the distance between the observed and forced monomer.

In summary, the coefficient k_{cn} can be estimated from Brownian simulations of a tethered polymer with self-avoiding interactions. The comparison with the experimental radius of confinement measured for chromatin loci [181, 193, 111, 39], reveals that the distance from the centromere is inversely correlated with the effective spring constant. Similar results are found for a tethered DNA molecule [158].

8.6.3 Polymer dynamics constricted by two tethering forces

The analysis of a single force can be generalized to the case of two monomers n and m ($n < m$) that are interacting with two distinct potential wells applied at positions $\boldsymbol{\mu}_n$ and $\boldsymbol{\mu}_m$ (Fig. 45b). The total potential energy of the Rouse polymer is

$$U_{\text{ext}}(\boldsymbol{R}) = \frac{1}{2}k_n(\boldsymbol{R}_n - \boldsymbol{\mu}_n)^2 + \frac{1}{2}k_m(\boldsymbol{R}_m - \boldsymbol{\mu}_m)^2, \tag{220}$$

and the average steady state position of the tagged monomer c is given by [17]

$$\langle \boldsymbol{R}_c \rangle = \begin{cases} \dfrac{\boldsymbol{\mu}_n k_n(\kappa + |m-c|k_m) + \boldsymbol{\mu}_m k_m(\kappa + |c-n|k_n)}{k_n k_m |m-n| + (k_n + k_m)\kappa} & n < c < m, \\[2ex] \dfrac{k_n \boldsymbol{\mu}_n \kappa + k_m \boldsymbol{\mu}_m(\kappa + |m-n|k_n)}{k_n k_m |m-n| + \kappa(k_n + k_m)} & n < m < c, \end{cases}$$

and

$$\lim_{\Delta t \to 0} \mathbb{E}\{\frac{\boldsymbol{R}_c(t+\Delta t) - \boldsymbol{R}_c(t)}{\Delta t} | \tilde{\boldsymbol{x}}\} = -D k_{cnm} \tilde{\boldsymbol{x}}, \tag{221}$$

where $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \langle \boldsymbol{R}_c \rangle$. The effective spring coefficient is given by [17]

$$k_{cnm} = \begin{cases} k_{cn} + k_{cm}, \text{ for } n < c < m \\[2ex] \dfrac{(2\kappa + |m-n|k)k\kappa}{\kappa^2 + |2c-m-n|k\kappa + |(m-n)(c-m)|k^2}, \text{ for } n < m < c \end{cases}$$

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where k_{cn} and k_{cm} are given by eq.218. For $n < m < c$, in the limit $m-n \gg 1$, then $k_{cnm} \sim |c-m|^{-1}\kappa$. The spring coefficient depends on the distance to the anchoring point only. When $n < c < m$, the effective spring coefficient depends on the positions of the two wells [279]. Finally, the variance of the monomer position with respect to its mean position (eq.221) is $\langle(\boldsymbol{R}_c - \langle\boldsymbol{R}_c\rangle)^2\rangle = \frac{d}{k_{cnm}}$. Thus, the distance scanned by the observed monomer is proportional to the distance to the anchoring point. (see fig.45b-ii,iii).

8.6.4 Analyzing tethering forces for a β-polymer

The β-polymer model [10] introduced in section 3.5, captures long-range interactions between monomers, that decay with the distance along the chain 3.5. Chromatin dynamics with an anomalous exponent less than 1/2 can thus be described by β-polymer models. When a gradient force (see eq. 214) is applied on monomer $\boldsymbol{R}_n$ of a β-polymer, the expectation of the velocity of monomer c ($c > n$) is given by [17]:

$$\lim_{\Delta t \to 0} \mathbb{E}\{\frac{\boldsymbol{R}_c(t+\Delta t) - \boldsymbol{R}_c(t)}{\Delta t} | \boldsymbol{R}_c(t) = \boldsymbol{x}\} = -Dk_{cn}(\beta, N, l, m)\boldsymbol{x}, \tag{222}$$

where $\boldsymbol{\mu} = 0$ and

$$k_{cn}(\beta, N, l, m) = A_{c,c} - \sum_{l,m \neq c} A_{l,c} A_{m,c} \tilde{C}^{-1}_{l,k}, \tag{223}$$

where $\tilde{C}$ is a block matrix, the i-th block of which is

$$\tilde{C}^i_{j,k} = A^i_{j,k} + k\delta_{i,n}\delta_{j,n}, \tag{224}$$

and [10]

$$A_{j,k} = \sum_{p=0}^{N-1} \tilde{\kappa}_p \alpha^j_p \alpha^k_p. \tag{225}$$

where the matrix α^j_p is defined in18. The inversion formula 222 for a β-polymer is similar to the one derived for a Rouse polymer (eq. 218), but the dependency with the parameters is now implicit [17]. Computing numerically the parameters in equations 223-225 reveals that the apparent spring constant $k_{cn}(\beta, N, l, m)$ decays slower with the distance $|c-n|$ (between the interacting and the observed monomer) and for smaller β (Fig. 45c). In summary, when the anomalous exponent is $\alpha < 0.5$, monomer trajectories are influenced by both the interacting monomers and the anomalous dynamics, characterized by the exponent α, that reflects the intrinsic chromatin organization.

8.6.5 Extracting forces acting on a locus from SPTs

We now present the empirical estimator for computing the effective spring coefficient k_c (eq. 218) from locus trajectories $\boldsymbol{R}_c(h\Delta t)$ ($h = 1..N_p$). We define the following empirical estimator

$$k_c \approx \frac{1}{d(N_p - 1)} \sum_{i=1}^{d} \sum_{h=1}^{N_p - 1} \frac{R^i_c((h+1)\Delta t) - R^i_c(h\Delta t)}{D\Delta t(R^i_c(h\Delta t) - \langle R^i_c\rangle)}, \tag{226}$$

where N_p is the number of points. The first step is to extract the mean position of the locus. Once the steady state is reached, the time average of the locus position is $\langle\boldsymbol{R}_c\rangle \approx \frac{1}{N_p}\sum_{h=1}^{N_p} \boldsymbol{R}_c(h\Delta t)$. This

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

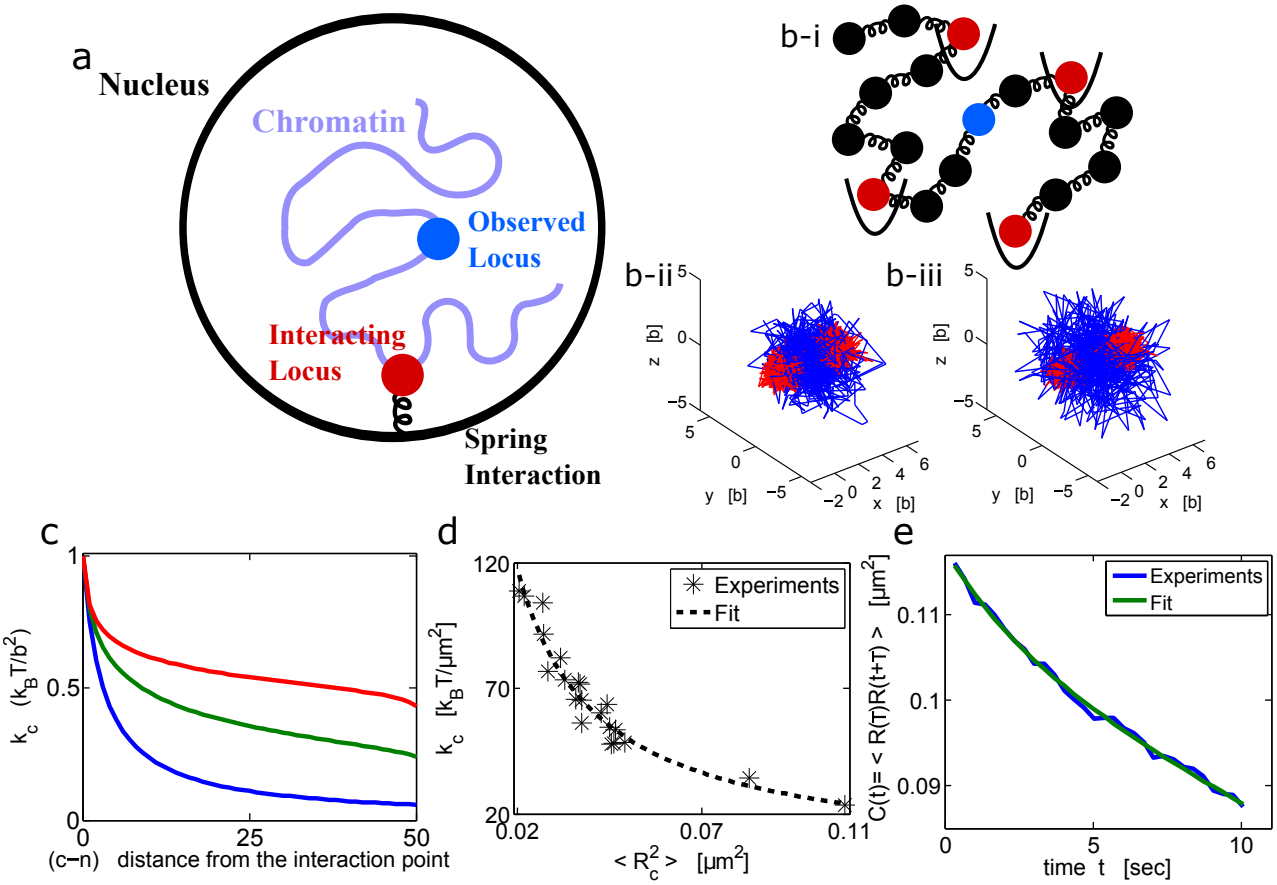


Fig. 45: **Dynamics of an observed locus when another locus is interacting with a nuclear element.** (**a**) Schematic representation of the nucleus, where one locus is observed and followed with a florescent label while another (non-visible) chromatin locus is interacting with another nuclear element. (**b-i**) Schematic representation of a polymer, where some monomers (red) interact with fixed harmonic potential wells. Monomer c (blue) is observed. (**b:ii,iii**) Stochastic trajectories of three monomers, part of a polymer where the two extremities interact with potential wells fixed at the origin and at position $\boldsymbol{\mu} = (5b, 0, 0)$ respectively. The middle monomer trajectory (blue) is more extended than the two others, shown for a polymer of length $N = 21$ (ii) and $N = 41$ (iii). (**c**) The apparent spring constant k_c is computed from formulas 218 and 222, for a polymer of length $N = 100$, where monomer $n = 50$ interacts with an harmonic potential (eq.214) with $k = 2k_BT/b^2$, while $\kappa = 3k_BT/b^2$. k_c is computed for increasing distances $|c - n|$, between the observed and the interacting monomers for $\beta = 2$ (Rouse polymer) (blue), $\beta = 1.5$ (green) and $\beta = 1.2$ (red). (**d-e**)Analysis of trajectories of the chromatin MAT-locus (located on chromosome III) in the yeast Sc. (**c**) Scatter plot of the effective spring coefficient k_c and the variance (R_c^2) of the locus trajectory estimated in two-dimensions, extracted for 21 cells. The constant k_c is estimated using formula 226, fitted to a power law, $k_c = \frac{a}{\langle R_c^2 \rangle^b}$, with $a = 3.03 \pm 1.05\, k_BT$ and $b = 0.94 \pm 0.1$. (**d**) Auto-correlation function computed using formula 191 for the trajectory shown in a. The fit uses the sum of two exponentials: $C(t) = a_1 e^{-t/\tau_1} + a_2 e^{-t/\tau_2}$, with $\tau_1 = 45.7 \pm 0.005s$ and $\tau_2 = 2.4 \pm 0.35$ s, $a_1 = 109 \pm 5 \times 10^{-3} \mu m^2$, $a_2 = 8.38 \pm 4.94 \times 10^{-3} \mu m^2$ (reproduced from [16]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

estimator can also be used when the polymer is not in equilibrium, but in this case $\langle R_c^i \rangle$ should be estimated separately. The diffusion coefficient can also be estimated using eq.202.

While k_c is estimated using eq.226, there are two unknown parameters: the spring force k and the distance $|c-n|$. For a strong anchoring ($k \gg \kappa$) and the approximation $k_c \approx \kappa |c-n|^{-1}$ is valid. The empirical effective spring constant can be used to estimate the distance to the interaction monomer.

For a long enough sampled trajectory and a force derived from a stationary potential well, the effective spring coefficient can be recovered directly either from the empirical estimator 226 or by using the reciprocal of the variance (eq.219). However, trajectories are often measured with a small sampling time Δt allowing probing the fine behavior of the chromatin and recovering accurately the diffusion coefficient. The total length of a trajectory is however limited by photobleaching effects [217]. Thus, the length of a trajectory may be shorter than the equilibration time scale, and thus acquired before equilibrium is reached. In that case, formula 226 can still be applied to recover the parameter k_c, while formula eq.219, which implies equilibrium, cannot be used.

8.6.6 Recovering applied forces from the auto-correlation function

The auto-correlation function $C(c, t_1, t_2) = \langle [\boldsymbol{R}_c(t_1) - \langle \boldsymbol{R}_c(t_1) \rangle] [\boldsymbol{R}_c(t_2) - \langle \boldsymbol{R}_c(t_2) \rangle] \rangle$ of a tagged monomer c is changed when a force is applied. Indeed, we decompose the energy potential of a polymer into the sum of internal Rouse plus an external component (eq.214) in the eigenvalue basis $\boldsymbol{u}_p$ that diagonalizes the Rouse potential 18. The external energy is

$$U_{\text{ext}}(\boldsymbol{R}_n) = \frac{1}{2} k \left(\boldsymbol{\mu} - \sum_{p=0}^{N-1} \alpha_p^n \boldsymbol{u}_p \right)^2, \tag{227}$$

and the Langevin equations for the polymer are given by

$$\frac{d\boldsymbol{u}_p}{dt} = D \left(k \alpha_p^n \boldsymbol{\mu} - ((\alpha_p^n)^2 k + \tilde{\kappa}_p) \boldsymbol{u}_p \right) - D k \alpha_p^n \sum_{q=0, q \neq p}^{N-1} \alpha_q^n \boldsymbol{u}_q + \sqrt{2D} \frac{d\widetilde{\boldsymbol{w}}_p}{dt}. \tag{228}$$

for $p = 0..N-1$. When the strength of the coupling term is relatively weak $(\alpha_p^n)^2 k \ll \tilde{\kappa}_p$, the coupling term can be neglected as well as higher modes for $k < \kappa$ and N large. Thus the auto-correlation function is

$$C(c, t_1, t_2) = \frac{d}{k} e^{-D(\alpha_0^n)^2 k (t_2 - t_1)} + \sum_{p=1}^{N-1} \frac{d(\alpha_p^c)^2}{(\alpha_p^n)^2 k + \tilde{\kappa}_p} e^{-D((\alpha_p^n)^2 k + \tilde{\kappa}_p)(t_2 - t_1)}. \tag{229}$$

The long-time asymptotic exponential decay can be used to extract the spring constant k of the applied force, when the diffusion coefficient D is known.

8.6.7 Distribution of forces of MAT-locus from live cell imaging SPT in yeast

We now discuss the implication of formula 226 to study the dynamics of a moving chromatin locus. We will present data from the MAT-locus acquired during 100sec with a time resolution of $\Delta t = 0.33$sec. Trajectories are covering a certain portion of the nucleus. Several predictions have been made based on formula 226: the computations on 20 cells revealed a large heterogeneity of k_c (Fig. 45d), where $\langle k_c \rangle = 67 \pm 22 k_B T / \mu\text{m}^2$. This heterogeneity indicates that in different cells the locus and the local chromatin interact differently with other nuclear elements. Plotting the force constant k_c for each

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cell with respect to the locus position averaged over the trajectory shows that ($\langle R_c^2 \rangle$) satisfies relation 218 (Fig. 45d), confirming that chromatin locus localization is governed by local interactions. The relation between k_c and $\langle R_c^2 \rangle$ is found empirically to be

$$k_c = \frac{3.0}{\langle R_c^2 \rangle^{0.94}}. \tag{230}$$

In addition, the auto-correlation function [237] for the MAT-locus position is computed from the estimator

$$C(t) = \frac{1}{N_p - t} \sum_{i=1}^{d} \sum_{h=1}^{N_p - t} R_c^i(h\Delta t) R_c^i((h+t)\Delta t), \tag{231}$$

and it is represented in fig.45e. Fractional Brownian motion has been previously used to model the dynamics of chromatin loci [138, 285, 127]. For fractional Brownian motion, the auto-correlation function $C(t)$ decays as a sum of power laws [130]. Thus the empirical correlation function $C(t)$ cannot be fitted by a power law, suggesting that the description of the locus motion as a fractional Brownian motion alone is not sufficient. However, a good fitting of the function $C(t)$ is obtained using a sum of two exponentials

$$C(t) \approx a_1 e^{-t/\tau_1} + a_2 e^{-t/\tau_2},$$

with $\tau_1 = 45.7 \pm 0.005$sec, $\tau_2 = 2.4 \pm 0.35$sec and $a_1 = 109 \pm 5 \times 10^{-3} \mu\text{m}^2$, $a_2 = 8.38 \pm 4.94 \times 10^{-3} \mu\text{m}^2$. This fit suggests that the auto-correlation function for the locus position is well described by a sum of exponentials, as predicted by formula 229 derived for general polymer model.

In summary, polymer models such as classical Rouse or $\beta-$ polymer accounts for the dynamics of a chromatin locus. In that context, it is possible to extract characteristics of the DNA locus, its dynamics, external forces and some properties of the polymer model from SPTs. At this stage, the origin of the anchoring force is not clear and may occur at the centromere. However, the large heterogeneity of the spring constant k_c characterizing an external applied force suggests (Fig. 45d) that the nearest anchoring points of that force is different for the same locus in different cells. Finally, the motion of the chromatin is driven by both thermal fluctuations and by active ATP-dependent forces [295]. While the above analysis is relevant to extract an interaction that does not change during the time acquisition of the trajectory, the spring constant k_c that would be extracted during an active chromatin motion could be differentiated from the thermal one by projecting the dynamics perpendicular to the direction of motion. Extracting active force elements may perhaps be done by considering the random force spectrum [90] or through the analysis of the correlation function, which are influenced by external forces.

9 Physical principles underlying DNA Repair

The nucleus organization can be revealed by the dynamics of loci at specific locations, such as nuclear pores at the membrane periphery, transcription regulation hot spots [184, 75, 32, 267], DNA repair foci [70], transient foci of proteins [51] or chromatin loci [244]. The life-time of these transient structures certainly determines their efficiency. Although the formation of these structures is still unknown, the search of homology during DNA repair is of particular interest and it consists of a random search by the broken strand for a homologous DNA template to repair the break. This section is dedicated to the polymer modelings describing the successive steps involved in the search of small targets and in particular during homologous recombination (HR).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

9.1 Encounter loci at the nuclear envelope

Although the role of the nuclear envelope is to separate chromosome from the rest of the cytoplasm, in differentiated nuclei, part of the chromatin (heterochromatin) is relocated along the inner face of the nuclear membrane, while open chromatin is found near nuclear pores. Evidence from yeast suggests that the nuclear envelope controls also gene transcription, DNA break repair and differentiation [3].

Yeast telomeres interact with the nuclear membrane through two redundant and cell cycle regulated pathways, involving telomere bound proteins Sir4 and Ku [265]. The perinuclear localization of this telomere can be increased upon activation of the HXK1 gene located 15kb upstream of the telomeric repeats. Transcriptional activation of the HXK1 in the presence of galactose increases its interaction with the nuclear pore complex and hence its perinuclear localization [266, 263] (Fig. 46a). In the presence of glucose, this gene becomes silent and on average it is relocated further away from the periphery. The localization of this particular telomere is mediated by the yKu protein in the G1 phase of the cell cycle. Interestingly, tagging a locus located near the subtelomeric region is used to visualize its reduced mobility (Fig. 46b), shown by the decrease of the length L_C (Fig. 46c). This reduction results from the interaction between the telomere and the nuclear periphery, confirmed by the increase of the coefficient k_c (Fig.46d). The diffusion coefficient is also reduced upon activation [266, 18].

To conclude, gene regulation can involve nuclear positioning and the spatial position of some genes is correlated with the expression level. In addition, some actively transcribed genes (Gal1-10, Ino1) contact the nuclear envelope through interactions with the nuclear pore complexes [263].

9.2 Physical constraints underlying double-stranded DNA break repair

Double-stranded DNA breaks (DSBs) can be caused by internal and external agents. Ionizing radiations lead to DSBs, while within the nucleus or promiscuous enzymes can cut both DNA strands. Cleavage of the DNA does not necessary implies that the two strands will separate [112]. In most cases, the two ends remain in close proximity with the assistance of repair proteins such as the MRN complex [93]. DSBs are the most serious and dangerous genetic lesions and if left un-repaired or repaired inaccurately, they can cause chromosomal translocations, cell death, loss of heterozygosity and can lead to the development of a cancerous cell.

DNA double-strand breaks can be repaired by one of two major classes of mechanisms: either by Non-Homologous End Joining (NHEJ), whereby the broken ends of the DNA are simply religated [93], or by homologous recombination (HR), which entails a physical search for a homologous DNA template by the broken strands, which is then used to repair the break [21, 112]. The template for repair may be the chromosomal homologue if the cell is diploid, or the sister chromatid if the break occurs after DNA replication. When a broken and resected DNA end comes close to the appropriate double-strand sequence, it invades the undamaged template and primes DNA synthesis. A subsequent mechanism called Holliday junction resolution leads to an accurate repair of the broken strand [102, 262, 2].

The search for the template is the rate-limiting for HR [182, 2, 289], but the exact mechanism is still unclear. It involves several spatial-temporal scales, from the molecular to the nuclear level, which are starting to be resolved experimentally and theoretically [88, 181, 69]. Several physical scenarios are possible to estimate the search time of the HR. The encounter time depends on the chromosome length, because the probability distribution of loci positions are determined by their length and distance from the centromere. The chromosome length can thus affect the probability of encounter: shorter chromosomes would tend to repair more amongst themselves, as would long chromosomes [2]. How can a single chromosomal locus find a copy of itself? How does it explore the territory of

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

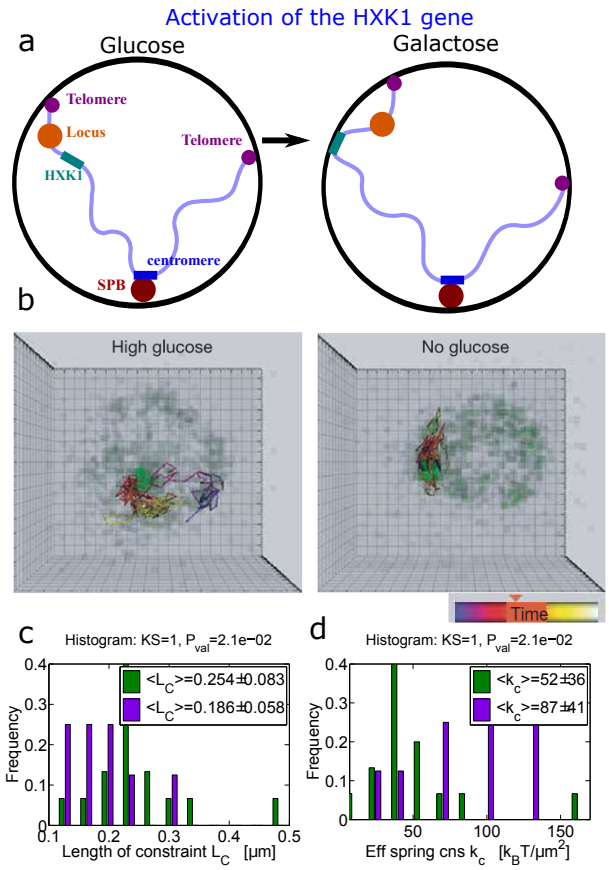


Fig. 46: **Transcription is regulated by gene re-positioning.** (**a**) The sub-telomeric locus ARS609 next to the right hand of telomere VI (TelVIR) in different growth conditions is compared: glucose (green) and galactose (purple). In both conditions the cells are in G1 [18]. (**b**) Four-dimensional time-lapse imaging of wild-type cells bearing lacop-tagged HXK1 as described [86] on glucose or galactose-containing media. Image stacks were collected at 1.5-s intervals for 7.5 min and tracking was performed with the Imaris Time program. The darker grey background shows the peripheral pore signal. The path of the locus is shown with a temporal color code (blue represents early and white represents late time points) [266]. (**c**) Distribution of the constrain length (L_C) found using eq.201. (**d**) Distribution of the apparent force strength (k_c) histogram for all cells as measured on the visible monomer c using eq.226.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

another chromosome in order to find a small sequence target [54]? The search for homology requires scanning of millions of base pairs for the correct sequence, but this process takes only between tens of minutes to a few hours [2, 49]. How is this possible? there is no final answer to that question. The surprising fast rate at which HR occurs is even more remarkable when considering that it is driven by diffusion and must overcome physical barriers and forces associated with the invasion of a chromatin [15].

The physical paradigm of HR is based on an experimental model for this search process: in these experiments, a well-characterized endonuclease mediated DSB at the MAT locus in haploid budding yeast cells can be induced by adding galactose (Fig. 47a). Following the expression of the endonuclease, the locus is cut with high efficiency. The locus is visualized by the binding of a fluorescent (LacI-GFP) protein at a Lac operator array (LacO), which is integrated in the chromosome at a distance 4.4kb from the cut site (Fig. 47a). Live-cell imaging studies in yeast have shown that a DSB scans a larger area of the nucleus than the unbroken locus, measured by the radius of confinement [181, 69, 15] (Fig. 47b). Equivalently, the value of the length of constraint (defined by eq.201), increases significantly after a break (Fig. 47c).

The DSB and the immediate surrounding chromatin can interact with the nuclear environment, the chromosomes, the nucleolus and known DSB anchored sites located at the nuclear envelope (nuclear pores and the SUN-domain protein Mps3 [265, 120]). The accumulated effect of these interactions impact the motion of a tagged locus. These interactions can be modeled as external forces acting on the monomer (discussed in section 8.6) and the strength of the force can be extracted from SPTs. From MAT locus trajectories, the effective spring coefficient k_c (formula 226) was shown to decrease significantly after break induction (Fig. 47d). This result shows that the chromatin structure around a DSB undergoes physical modifications, reflected by a change in the forces applied on the chromatin. This effect suggests that a tagged locus scans a larger nuclear area, emphasized by an increase in the length of constraint L_C (Fig. 47c). Following a DSB [258], the attachment of the centromere to the kinetochore is relieved, resulting in a reduction in k_c.

To conclude, the physical changes after break induction in yeast and engagement into the HR repair pathway, support the theory that the chromatin is locally reorganized, nucleosomes are first rapidly acetylated, H2A is phosphorylated, and then evicted around the break, coincident with end-resection [276].

9.3 Increasing DSB dynamics reduces the MFET

What is the consequence of reducing local interactions following a DSB motion? it might change the HR search time. Indeed, during HR repair of a DSB, chromatin structure may expand [252], facilitating both the search process [208], and the binding of the repair machinery [298]. Yet, how the chromatin structure is modified to favor this search process is unclear [203, 98]. One possible scenario is the release of forces associated with the relaxation/decondensation of chromatin. This expansion of the chromatin was characterized using the $\beta-$polymer model. The anomalous exponent α increases following break induction (Fig. 47f). This increase reflects a change in the local property of the polymer model (the chromatin): when it is small ($\beta \to 1$), all monomers are highly interacting and the polymer has a compacted shape, while for large values of β ($\beta \to 2$) the polymer is more relaxed (Fig. 47g). After DSB induction β increases, leading to a wider distribution of locus (monomer) position. The increase of the anomalous exponent α could also be due to active mechanisms as described in section 8.4.3.

Brownian simulations of polymer models suggest [15] that chromatin relaxation accelerates the search for homology by factor 4 to 8, by increasing the locus mobility and reducing the screening

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

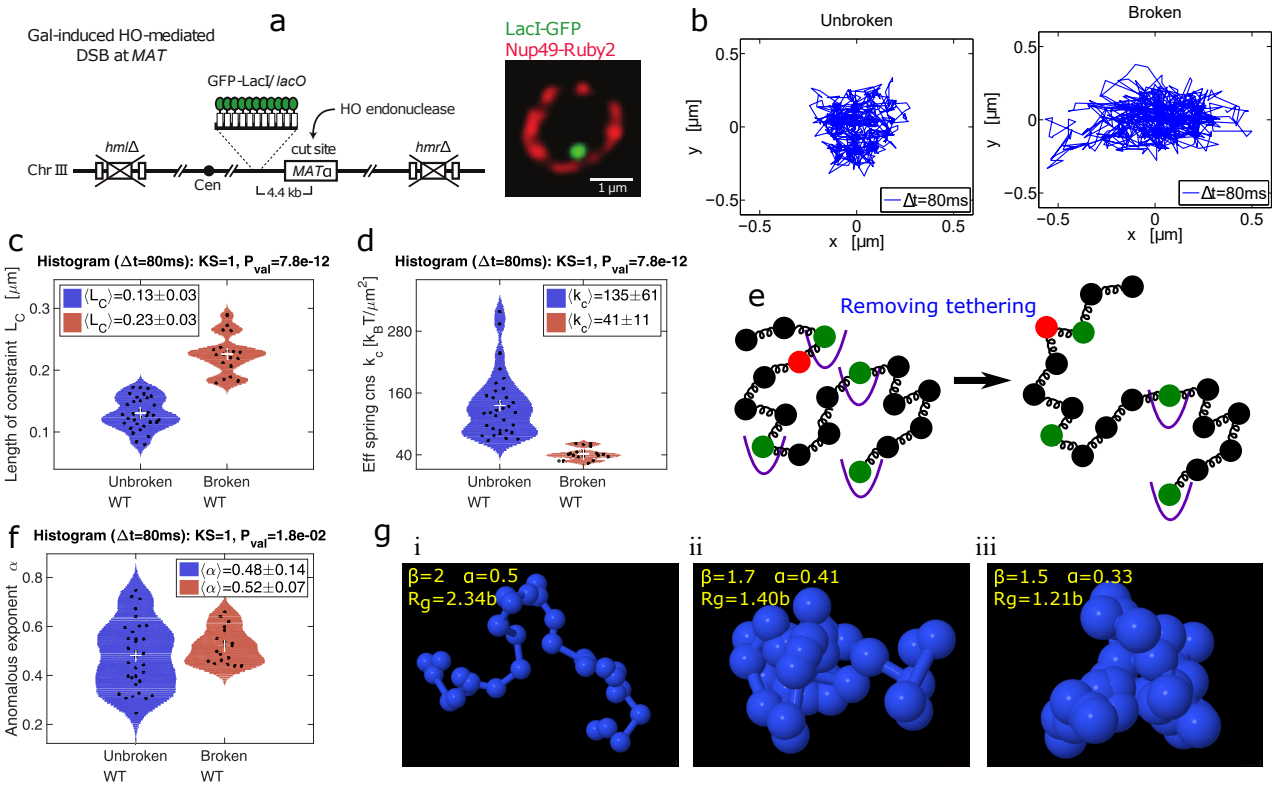


Fig. 47: **The dynamics of a double stranded DNA break in the MAT locus.** **(a)** Schematics of the experimental system. **(b)** (left) The MAT locus trajectory during 60 seconds (Δt = 80ms). (right) The MAT Locus trajectories 2 hours after the induction of a break, during repair with the homologous recombination pathway. **(c)** The length of constraint calculated from the trajectories using eq.201 of a broken locus (red) vs. the unbroken locus (blue). In the title we present the result of a Kolmogorov-Smirnov and the P value. **(d)** Quantification of the effective spring coefficient k_c (eq.226) before and after break induction. **(e)** Loss of chromatin sites interactions: Schematic representation of two homologous polymers, where a fraction of monomers (green) participate in a steady state external interaction, modeled by a fixed harmonic potential wells (purple), while single trajectories are recorded from monomer c (red). **(f)** The anomalous exponent α is extracted from the cross-correlation function of trajectories for a broken (red) vs. unbroken (blue) locus, with time steps of 80msec. **(g)** Three examples of β-polymers with **(a)** $\beta = 2$, **(b)** $\beta = 1.7$ and **(c)** $\beta = 1.5$, where the inter-monomer interactions are modified, resulting in different condensation state. The radius of gyration R_g (red) measures the degree of compaction associated with β, which is related the to anomalous exponent of a monomer by the relation $\alpha = 1 - 1/\beta$. Balls (blue) represent the monomers of radius $0.3b$. After break induction α increase, which indicates the chromatin expansion (reproduced from [15]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

barrier, imposed when the broken site remains hidden in the chromatin structure. When a monomer stays inside the chromatin domain, the surrounding monomers generated expulsion forces between themselves and any other monomers belonging to another polymer (see also subsection 4.12). When a monomer located inside a chromatin domain needs to get in close proximity to a template site, modeled as a monomer inside another polymer, it would have to penetrate into that domain. The interactions between the search monomer and monomers surrounding the target generate an effective potential barrier, that the searcher needs to overcome in order to encounter its target. This potential is due to the local effect of all interacting monomers (see Fig. 4 in [8]). However, when a monomer is robustly positioned at the periphery of a polymer globule, when the chromatin is open, the effective potential is reduced, accelerating the search time [15]. This situation is similar to the scenario of classical chemical reaction where a catalyst reduces the activation potential barrier, accelerating the reaction. It would be possible to use polymer models to estimate the timescales of relocation a DSB toward the nuclear envelope, although precise information would be needed about the chromatin condensation for various distances to the boundary.

In summary, releasing tethering forces around a DSB enhances its motion and significantly reduces the homology search time. Thus, the DSB interaction capacity (the external forces applied to the DSB region) with other chromosomes is reduced, allowing an efficient search for a template.

9.4 DSB re-localization as a protection mechanism

In budding yeast persistent DBSs that do not have a homologous donor migrate to the nuclear envelope [188, 120] (Fig. 48a), where they are anchored either at nuclear pores or at the SUN domain protein, Mps3. The relocalization of the HO-induced DSB to the nuclear envelope is not immediate, occurring between 30 minutes and two hours after cut induction, yet it persists for over four hours. This re-localization can serve as a protection from translocation with other chromatin sites.

The changes in the motion of a DSB after break induction are characterized by an increase in the diffusion coefficient, the anomalous exponent and the confinement radius, confirming physical changes at the chromatin site (Fig. 47b-c-d-f). These changes depend on repair proteins [181, 69, 15] and can be caused by nucleosome repositioning [15]. In some cases, the broken site is extruded from the interior of a dense chromatin territory, and it relocates to the periphery of the local chromatin domain (Fig. 48b). This reconfiguration is due to a reduction of the internal chromatin constraints and result in the extrusion of the DSB. Consequently, it may facilitate sequence-sequence encounters in a condensed chromatin environment. Several observations support this concept: the homologous recombination protein Rad51 is recruited to and forms foci at DSBs. Because heterochromatin may have repetitive sequences, there is a danger that repairing by NHEJ with the domain will lead to chromosome translocation. Thus, these foci move from the interior of the heterochromatin domain to the periphery to complete recombination [49] (Fig. 48c). Similarly, in *S. cerevisiae*, to form a Rad52 focus on broken rDNA, the locus must move out of the nucleolus [276] (Fig. 48d). Rad51 dependent telomere movement was involved in the alternative lengthening of telomeres (ALT) [50]. These modifications depend on the activity of repair proteins Rad51 and Rad54 which facilitate the extrusion of the DSB. A deletion of either one decreases the anomalous exponent α, allowing a strong increase in local intrinsic chromatin tethering forces [15]. In the absence of Rad51 or Rad54, a DSB becomes more constrained.

In summary, local condensation or compaction of nucleosomes around the break site is due to an increase in local tethering forces. This effect may either be a protective mechanism to avoid translocations or it can influence the microhomology mediated end-joining pathway [62]. If HR is

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

not functional, the damage is sequestered in a compact nanodomain, possibly to facilitate end-joining. To conclude, in a crowded environment such as the nucleus, self-avoiding interactions have a dominant effect on the encounter time and probability between two chromatin locus. These interactions can be regulated in numerical simulations by changing the monomer density and dynamics and is mediated experimentally by an ensemble of chromatin modification proteins during repair or transcription.

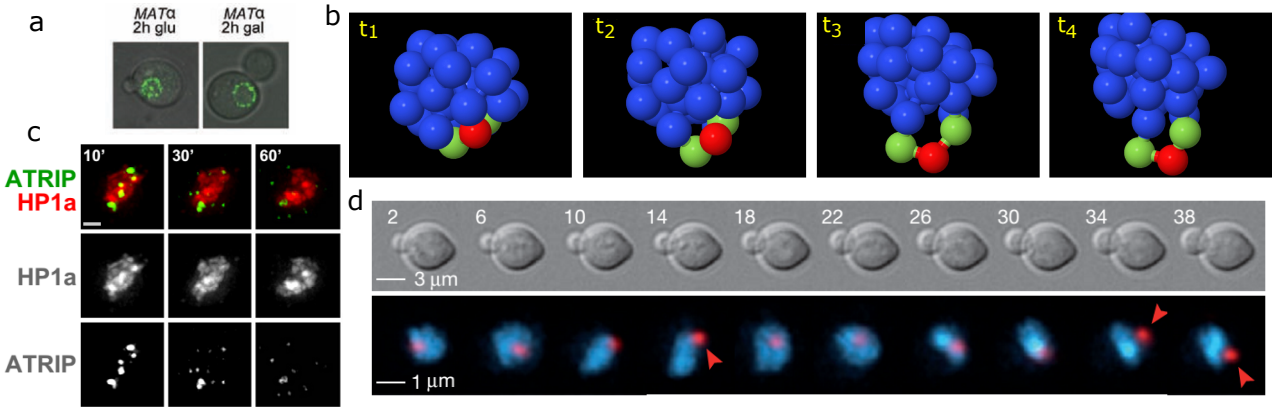


Fig. 48: **DSB motion can serve to increase repair efficiency and reduce the change for chromosomal translocations.** (a) Un-repairable DSB move to the periphery: When a DSB cannot repair itself for a long time, it will be anchored an the nuclear pores. Before break induction (growth in glucose), the locus is randomly distributed in the nucleus. The MAT locus relocate to the periphery between 0.5 and 2 hours after the HO-induced DSB (using galactose) [188]. (b) Upon the induction of a DSB in condense chromatin domain, the DSB is gradually extruded to the periphery of a domain. The process is modeled using a β polymer, where at t_1, the long-range connections for monomers $n = 16, 17, 18$ (in a chain of length $N = 33$) are cut. As a result monomer $n = 17$ is slowly pushed ($t = t_2, t_3, t_4$) to the periphery of the polymer globule. (c) DSB move outside of heterochromatin after irradiation of Drosophila nuclei [49]: ATRIP foci (ATRIP is a repair protein which is recruited to the DSB) form within the HP1a (heterochromatin/condense chromatin) domain and relocate outside of this domain. Stills images of cells expressing GFP-ATRIP and mCharry-HP1 (HP1 is a heterochromatin associated protein) are shown, in which DSBs were produced by ionizing radiation (IR). Maximum intensity projections of one cell at 10, 30, and 60 min after IR show the relocalization of ATRIP foci to the periphery and outside of the HP1a domain. (d) A DSB in rDNA locus is moved outside of the nucleolus [276]: Time-lapse microscopy of a TetImRFP-focus (red) next to a DSB marked rDNA repeat in a yeast cell expressing Nop1CFP (blue region). Arrowheads mark the exit of the marked rDNA repeat with the DSB from the nucleolus. Time is indicated in minutes. Selected time points of a representative cell are shown (reproduced from [15]).

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10 Conclusion and perspective

Polymer models have a significant role in clarifying chromatin and nucleus organization. These models are used for extracting biophysical parameters and interpreting single particle trajectories of a chromatin locus. They further serve for reconstructing the chromatin organization from large data sets such as chromosomal capture data. Refined modeling, coarse-grained stochastic simulations, analysis of the models are now possible for major nuclear events such as double stranded DNA repair and differentiation, but also for the smaller scale function of a cell involving transcription, gene activation or nucleosome translocation. We expect much progress in the next decade in integrating three dimensional super-resolution data with a nanometer precision into polymer models. Indeed, with the improvement of real-time, high resolution imaging, we shall be able to study the manner at which proteins interact with the chromatin.

DNA and chromatin spatial scales span one to two orders of magnitude between bacteria and mammalian cells: while yeast chromosomal organization is fairly well described by classical polymer models based on anchoring, self-avoiding interaction and confinement, metazoans chromosome structure would require more sophisticated models. It remains unclear how to interpret numerical simulations of polymers reconstructed from Hi-C data, because these data represent an average of many realizations, where geometrical distances to the nucleus surface are disregarded. As discussed here, the interpretation of chromosome capture experimental results and their finer details remain unclear. These data represent a frozen conformation of a cell and it is not clear how far they have distorted in-vivo configurations. Developing novel analytical and algorithm methods to reconstruct the chromatin organization should certainly continue clarifying the local chromatin re-organization. Understanding and quantifying the nucleus dynamics requires also novel development in the asymptotic analysis of the polymer models, to extract relations between parameters. These formula will be a fundamental tool to summarize the laws of epigenetic-physics.

10.1 Which polymer model to use and when?

Today, the overall space of polymer physics is certainly not completely understood or described. But the increasing need for interpreting data has push to chose a correct (or not) appropriate polymer model for any given situation. We shall now provide some indications about which polymer model should be used to analyse data. Indeed, several statistical quantities can be analytically computed, large and heavy simulations are possible but they all depend on a chose of a polymer model so which one to chose?

In general, there are no consensus about using one polymer model rather than another. For example, the Hi-C data remains difficult to analyse and as discussed in sections 5.4.3, but the compact chromatin is taken into account by the string binder and switch model [19, 20] using random loops and binding sites. Recently, the binder model was extended to impose link between monomers that are highly connected in the Hi-C matrix [246], so that the decay of the encounter probability matrix is recovered. The method allows to prescribed the number of random connectors positioned to reconstruct TAD.

The classical Rouse model was used to interpret single particle trajectories where the anomalous exponent of a given Chromatin locus is $\alpha = 0.5$. For bacteria, where the exponent α is distributed with a mean at 0.4 [128], another possibility is to use $\beta-$ polymer model, discussed in section 5.4.2 to generate a polymer model with a prescribed anomalous exponent ($\alpha \leq 0.5$). However the interpretation of exponents $\alpha \geq 0.5$ remains difficult as it certainly involved a mixture of diffusion and active transport. For modeling telomere dynamics in yeast, polymer models were based on spring

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

forces and repulsion force between beads [291] (see also section 7). In all cases, polymer model can be used to obtain statistics about transient events and reconstruct the histogram of arrival time between any two sites of interest (see section 6.4). Another use of polymer model is to study some properties of the search of homologous side during HR (see section 9.3). Classical Rouse or with some modifications (beta-polymer or with repulsion) have been used to study the influence of the nucleus membrane on the the search.

In the context of interpreting single locus trajectories, polymer models serve for predicting the statistics to be expected from data such as the anomalous exponent, radius of confinement, etc.. (summarized in section 8). But the statistics are not necessarily polymer model specific. For example, the first and second moments (see formula 221) of the velocity can always be computed, but their interpretation require a specific polymer model.

10.2 Possible directions and open questions

We end this section with some open directions where modeling and experimental approaches that should combine:

10.2.1 Extracting forces from multiples tagged loci

What can be extracted from several single locus trajectories. For example how use the properties of the correlation function? the goal would be to extract intrinsic mechanical properties of the chromatin and applied forces. This question is now accessible due to the possibility of tagging at least two loci with a known distance [65, 108]. The theory remains to be worked out and could be based on generalizing eq. 221 to two known locus positions. The cross-correlation function would provide information the rigidity of the interaction. The difficulty is to get long enough trajectories.

10.2.2 Is there a best model to describe Hi-C data and the encounter probability matrix?

What should be the correct model to describe Hi-C data. As discuss in the previous subsection, there are a few models [19, 20, 91, 196]. Possible improvements may by based on using simultaneously random loops and local peaks in the data distribution [246]. It would be also interesting to reconstruct the polymer model by including the geometrical organization of euchromatin and heterochromatin and the nucleus membrane. No such data are yet available, but the application of this approach would be to have a precise location of specific gene in space. We expect that new finding is expected about the organization of gene activated together.

10.2.3 How to reconciliate SPTs and Hi-C statistics?

At the present moment, statistics of SPTs and Hi-C have not been really compared. However, polymer models can be used to compare the decay of the encounter probability of a given locus and the anomalous exponent of that locus. Both quantities are related for example through the $\beta-$ polymer model (see section 5.4.2). However, the deviation from the formula 71 (section 3.10.4) should reveal local multiple chromatin interactions. It should be possible to start with computing the anomalous exponent from numerical simulations.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10.2.4 Proteins remodeling during DNA break repair

Double-stranded DNA break repair involves many repair proteins at different stage of the process (see sections 9.2 and 9.4). For example Rad51 and Rad54 were shown to contribute to the increased movement of DSBs in yeast [68] and Rad51 might facilitate the extrusion of a DSB. Nucleosome reorganization is critical when repairing a break in heterochromatin where nucleosome remodelers invade and pair. In budding yeast, Arp8 is a key component of the INO80 nucleosome remodeling complex and is required for efficient recombination [2] and for increasing the mobility for DSBs and undamaged sites [242, 68]. How shall we model the effect of these proteins during DSB break repair? for example, how the local forces (spring constants) on a Rouse model should be locally modified and to what extend? The polymer model can be modified by considering multiple states for the local monomers near the break. It would be interesting to generate numerical polymer simulations of single locus and estimate various quantities of interested such the diffusion coefficient, interaction forces, the anomalous exponent and so on. A comparison with SPTs should reveal refined repair steps [15].

10.2.5 Revisiting the classical question of the search process for a promoter site by a transcription factor

With the unprecedent resolution of the chromatin structure and live cell organization, the classical search process scenario of a promoter site by a TF, studied by von Hippel and colleagues (see also [167] for an analytical approach) should be revisited. Indeed, it would be now possible to take into account the relative positions of binding sites to reconstruct the local chromatin environment using the encounter probability matrix of the Hi-C data. Super-resolution data would further refined the exact geometry for stochastic simulations to explore how the chromatin organization can modulate the stochastic search process and in particular, what is really the time spends in three dimension versus the time spend scanning along the DNA.

10.2.6 Computing the probability of a translocation during Non-Homologous End-Joining (NHEJ)

We summarize here one aspect of repair, called HR (setion 9). However, another repair pathway is the NHEJ where two broken end simply relegate. However, when there are several break, some can relegate incorrectly (with another broken arm). This process is call translocation. It would be important to estimate the probability of such translocation event. No much simulations or analysis have been pushed in that direction.

10.2.7 Modeling the random initiation of gene factories

The looping time between any two monomers is now clarified (see section 4), however the mean first time for k ($k \geq 2$) monomers of a long polymer to enter into a ball of radius ε has not yet been computed. This encounter may represent the process of initiating a gene factory regulation in a subdomain. Indeed, it may conceivable that several promoters and repressors interact in small regions to regulate gene expression [209]. The difficulty in the computation of the encounter time is due to additional constraint in the computations of the Gaussian integrals (see relation 99). It would be important to derive analytical expressions as computer simulations of this process should converge slowly due to the singular perturbation: indeed, the encounter time diverges to infinity when the radius ε tends to zero.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10.2.8 Nucleosome repositioning following DNA damages

Nucleosomes are repositioned following DNA damages, but there is no model for this changes because both the chromatin and the nucleosome are changing their organization. It would be to difficult to account for the molecular detail of the nucleosomes, so a coarse-grained model is necessary. The difficult is chromatin and nucleosomes do not have te same spatial scale, which render the modeling and simulation difficult. Such simulations would help exploring how nucleosome and chromatin are reoganized following break and how the re-organized during repair.

10.2.9 Limitation of statistical analysis for High C-data

Hi-C data have been so far analyzed by estimating statistical parameters such as the decay law for the encounter probability predicted by polymer models [196]. It is however possible to go one step further and design realistic stochastic simulations of chromatin constraint on chromosomal High C-data. We are still facing the limitation of the data that only represent an average configuration over a large cell population, but single cell Hi-C is now possible. Simulations could be used to explore the probability and the conditional mean time for particular sites to meet and also to explore and classify the ensemble of polymer configurations (see also [246]for a first step).

10.2.10 Can statistical analysis of Hi-C data reveal novel features hidden in the genome organization of cancer cells or cells infected by viruses

We speculate that the nucleus organization in cancer cells might reveal clear difference with normal cells, although there no available data generated so far in that direction. Similarly, viral DNA inserted at specific random locations in the chromatin of the host cell might lead to changes in the encounter probability of the Hi-C data. SPT dynamics with polymer reconstruction might lead to a better understanding of the consequences host-virus interactions.

Acknowledgment

We would like to thank the anonymous reviewers for constructive remarks and A. Taddei, K. Dubrana, A. Seeber, S. Gasser, X. Darzacq, Z. Schuss, L. Atkinson, E.Laue, I. Kupka, J.M. Victor, E. Fabre, L. Mirny, Z. Schuss, Y. Kantor, C. Lavelle, P. Ciccuta, L. Georgetti, O. Shukron, R. Metzler, M. Barbi, E. Heard, A. Bancaud, I. Sokolov, S. Redner, M. Kardar, C. Lavelle, Y. Garini, C. Vaillant, I. Cisse, C. Zimmer and G. Almouznie for stimulating discussions over the past ten years. D.H. thanks the Mathematical Institute, University of Oxford for hospitality, this research was partially supported by a Simons fellowship from the Newton Institute.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, 1972.

[2] N. Agmon, Liefshitz B., Zimmer C., Fabre E., and Kupiec M. Effect of nuclear architecture on the efficiency of double-strand break repair. *Nat. Struct. Mol. Biol.*, 15(4):694–9, 2013.

[3] A. Akhtar and S. M. Gasser. The nuclear envelope and transcriptional control. *Nature reviews. Genetics*, 8(7):507–17, July 2007.

[4] B. Albert, J. Mathon, A. Shukla, H. Saad, C. Normand, I. Léger-Silvestre, D. Villa, A. Kamgoue, J. Mozziconacci, H. Wong, C. Zimmer, P. Bhargava, A. Bancaud, and O. Gadal. Systematic characterization of the conformation and dynamics of budding yeast chromosome XII. *Journal of Cell Biology*, 202(2):201–210, July 2013.

[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* Garland Science, New York, 2007.

[6] Elnaz Alipour and John F. Marko. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Research*, 40(22):11202–11212, 2012.

[7] J. Allemand, S. Cocco, N. Douarche, and G. Lia. Loops in dna: An overview of experimental and theoretical approaches. *Eur. Phys. J. E*, 19:293–302, 2006.

[8] A. Amitai, C. Amoruso, A. Ziskind, and D. Holcman. Encounter dynamics of a small target by a polymer diffusing in a confined domain. *J. Chem. Phys.*, 137:244906, 2012.

[9] A. Amitai and D. Holcman. Diffusing polymers in confined microdomains and estimation of chromosomal territory sizes from chromosome capture data. *Phys. Rev. Lett.*, 110:248105, 2013.

[10] A. Amitai and D. Holcman. Polymer model with long-range interactions: Analysis and applications to the chromatin structure. *Phys. Rev. E*, 88:052604, 2013.

[11] A. Amitai, Y. Kantor, and M. Kardar. First-passage distributions in a collective model of anomalous diffusion with tunable exponent. *Phys. Rev. E*, 81:011107, 2010.

[12] A. Amitai and D. Kupka, I.and Holcman. Analysis of the mean first looping time of a rod-polymer. *Multiscale Model. Simul.*, 10:612, 2012.

[13] A. Amitai, I. Kupka, and D. Holcman. Computation of the mean first-encounter time between the ends of a polymer chain. *Phys. Rev. Lett*, 109:108302, 2012.

[14] A. Amitai, I. Kupka, and D. Holcman. Kinetics of diffusing polymer encounter in confined cellular microdomains. *J. Stat. Phys.*, 153:1107–1131, 2013.

[15] A. Amitai, C. Seeber, S. M. Gasser, and D. Holcman. Visualization of chromatin decompaction and break site extrusion as predicted by statistical polymer modeling of single locus trajectories. *Cell Report*, 2017.

[16] A. Amitai, M. Toulouze, K. Dubrana, and D. Holcman. Analysis of single locus trajectories for extracting in vivo chromatin tethering interactions. *PLoS Comput Biol*, 11:e1004433, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[17] A. Amitai, M. Toulouze, K. Dubrana, and D. Holcman. Extracting in vivo interactions acting on the chromatin from a statistical analysis of single locus trajectories. *Personal Communication*, 2015.

[18] A. Amitai, M. Toulouze, K. Dubrana, and D. Holcman. Heterogeneity of dna-locus dynamics revealed by polymer physics analysis. *personal communication*, 2015.

[19] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences of the United States of America*, 109(40):16173–8, October 2012.

[20] M. Barbieri, J. Fraser, L.-M. Lavitas, M. Chotalia, J. Dostie, A. Pombo, and M. Nicodemi. A polymer model explains the complexity of large-scale chromatin folding. *Nucleus (Austin, Tex.)*, 4(August):267–273, 2013.

[21] A. Barzel and M. Kupiec. Finding a match: how do homologous sequences get together for recombination? *Nature Rev. Genet.*, 9:27–37, 2008.

[22] D. Baù, A. Sanyal, B. R. Lajoie, E. Capriotti, M. Byron, J. B. Lawrence, J. Dekker, and M. A. Marti-Renom. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nature structural & molecular biology*, 18(1):107–114, 2011.

[23] Fabrizio Benedetti, Julien Dorier, Yannis Burnier, and Andrzej Stasiak. Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. *Nucleic Acids Research*, 42(5):2848–2855, 2014.

[24] O. Bénichou, C. Chevalier, B. Meyer, and R. Voituriez. Facilitated diffusion of proteins on chromatin. *Physical Review Letters*, 106(January):1–4, 2011.

[25] A. B. Berger, G. G. Cabal, E. Fabre, T. Duong, H. Buc, U. Nehrbass, J. Olivo-Marin, O. Gadal, and C. Zimmer. High-resolution statistical mapping reveals gene territories in live yeast. *Nature methods*, 5(12):1031–7, December 2008.

[26] T. Bickel. A note on confined diffusion. *Physica A*, 377:2432, 2007.

[27] E. H. Blackburn. Structure and function of telomeres. *Nature*, 350(6319):569–73, 1991.

[28] Alistair N Boettiger, Bogdan Bintu, Jeffrey R Moffitt, Siyuan Wang, Brian J Beliveau, Geoffrey Fudenberg, Maxim Imakaev, Leonid A Mirny, Chao-ting Wu, and Xiaowei Zhuang. Chromatin Folding for Different Epigenetic States. *Nature*, 529(7586):1–15, 2016.

[29] M. Bohn, , D. W. Heermann, and R. van Driel. Random loop model for long polymers. *Phys. Rev. E*, 76:051805, 2007.

[30] M. Bohn and D. W. Heermann. Diffusion-driven looping provides a consistent provides a consistent framework for chromatin organization. *PLoS ONE*, 5(8), 2010.

[31] H. G. Bohr. *Handbook of molecular biophysics: methods and applications*. Wiley, VCH, 2009.

[32] M. R. Branco and A. Pombo. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS biology*, 4(5):e138, May 2006.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[33] D. Brockmann and T. Geisel. Particle dispersion on rapidly folding random heteropolymers. *Phys. Rev. Lett.*, 91:048303, 2003.

[34] I. Bronstein, Y. Israel, E. Kepten, S. Mai, Y. Shav-Tal, E. Barkai, and Y. Garini. Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Phys. Rev. Lett.*, 103:018102, Jul 2009.

[35] Robijn Bruinsma, Alexander Y Grosberg, Yitzhak Rabin, and Alexandra Zidovska. Chromatin hydrodynamics. *Biophysical journal*, 106(9):1871–81, may 2014.

[36] Yannis Burnier, Julien Dorier, and Andrzej Stasiak. DNA supercoiling inhibits DNA knotting. *Nucleic Acids Research*, 36(15):4956–4963, 2008.

[37] C. Bustamante, S. B. Smith, J. Liphardt, and Smith D. Single-molecule studies of dna mechanics. *Curr. Op. Struct. Biol.*, 10:279–285, 2000.

[38] O. V. Bychuk and B. O'Shaughnessy. Anomalous diffusion at liquid surfaces. *Phys. Rev. Lett.*, 74:1795, 1995.

[39] K. Bystricky, P. Heun, L. Gehlen, J. Langowski, and S. M. Gasser. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. U. S. A.*, 47:16495–16500, 2004.

[40] K. Bystricky, T. Laroche, G. van Houwe, M. Blaszczyk, and S. M Gasser. Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *J Cell Biol*, 168:375–87, 2005.

[41] K. Bystricky, H. Van Attikum, M. D. Montiel, V. Dion, L. Gehlen, and S. M. Gasser. Regulation of nuclear positioning and dynamics of the silent mating type loci by the yeast ku70/ku80 complex. *Mol Cell Biol*, 29:835–48, 2009.

[42] Luijten E. Cacciuto A. Self-avoiding flexible polymers under spherical confinement. *Nano Lett.*, 6:901–5, 2006.

[43] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasherf. Green Fluorescent Protein as a Marker for Gene Expression. *Science*, 263(1988):802, 1994.

[44] Tamir Chandra, Kristina Kirschner, Jean Yves Thuret, Benjamin D. Pope, Tyrone Ryba, Scott Newman, Kashif Ahmed, Shamith A. Samarajiwa, Rafik Salama, Thomas Carroll, Rory Stark, Rekin's S. Janky, Masako Narita, Lixiang Xue, Agustin Chicas, Sabrina N?nez, Ralf Janknecht, Yoko Hayashi-Takanaka, Michael D. Wilson, Aileen Marshall, Duncan T. Odom, M. Madan Babu, David P. Bazett-Jones, Simon Tavar, Paul A W Edwards, Scott W. Lowe, Hiroshi Kimura, David M. Gilbert, and Masashi Narita. Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Molecular Cell*, 47(2):203–214, 7 2012.

[45] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15:1–89, Jan 1943.

[46] I. Chavel and E. A. Feldman. Spectra of manifolds less a small domain. *J. Duke Math.*, 56:399–414, 1988.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[47] J. Z. Y. Chen, H. K. Tsao, and Y. J. Sheng. Diffusion-controlled first contact of the ends of a polymer: Crossover between two scaling regimes. *Phys. Rev. E*, 72(3):031804, 2005.

[48] A. F. Cheviakov and M. J. Ward. Optimizing the principal eigenvalue of the laplacian in a sphere with interior traps. *Math. Comput. Modelling*, 53:1394–1409, 2011.

[49] I. Chiolo, A. Minoda, S. U. Colmenares, A. Polyzos, S. V. Costes, and G. H. Karpen. Double-strand breaks in heterochromatin move outside of a dynamic hp1a domain to complete recombinational repair. *Cell*, 144:732–44, 2011.

[50] N. W. Cho, R. L. Dilley, M. A. Lampson, and R. A. Greenberg. Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell*, 159(1):108–21, September 2014.

[51] I. Cisse, I. Izeddin, S. Z. Causse, L. Boudarene, A. Senecal, L. Muresan, C. Dugast-darzacq, B. Hajj, M. Dahan, and X. Darzacq. Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science*, 341(2007):664–667, 2013.

[52] D. Corkery, K. L. Cann, and G. Dellaire. *Nuclear Domains, in Cellular Domains (ed I. R. Nabi)*. John Wiley and Sons, Inc., Hoboken, NJ, USA, 2011.

[53] Ruggero Cortini, Maria Barbi, Bertrand R Caré, Christophe Lavelle, Annick Lesne, Julien Mozziconacci, and Jean-Marc Victor. The physics of epigenetics. *arXiv preprint arXiv:1509.04145*, 2015.

[54] T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, 2:292–301, 2001.

[55] Holcman D. and Schuss Z. Control of flux by narrow passages and hidden targets in cellular biology. *Rep Prog Phys.*, 76(7):074601, 2013.

[56] F. de Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner, T. Lagache, A. Dufour, and J. Olivo-Marin. Icy: an open bioimage informatics platform for extended reproducible research. *Nature methods*, 9(7):690–6, July 2012.

[57] P. G. de Gennes. *Scaling Concepts in Polymer Physics*. Ithaca, New York: Cornell Univ. Press, 1979.

[58] P. Debye and E. Hückel. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. The theory of electrolytes. I. Lowering of freezing point and related phenomena. *Physikalische Zeitschrift*, 24:185–206, 1923.

[59] J. Dekker. Two ways to fold the genome during the cell cycle : insights obtained with chromosome conformation capture. *Epigenetics Chromatin*, 7(1):1–12, 2014.

[60] J. Dekker, M. A Marti-Renom, and L. A. Mirny. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics*, 14(6):390–403, 2013.

[61] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–11, 2002.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[62] S. K. Deng, B. Gibb, M. J. de Almeida, E. C. Greene, and L. S. Symington. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nature structural & molecular biology*, 21(4):405–12, April 2014.

[63] W. H. Deng and E. Barkai. Ergodic properties of fractional brownian-langevin motion. *Phys. Rev. E*, 79:011112, 2009.

[64] Edmund A Di Marzio and Arnold J Mandell. Phase transition behavior of a linear macromolecule threading a membrane. *The Journal of chemical physics*, 107(14):5510–5514, 1997.

[65] D. Dickerson, M. Gierlinski, V. Singh, E. Kitamura, G. Ball, T. Tanaka, and T. Owen-Hughes. High resolution imaging reveals heterogeneity in chromatin states between cells that is not inherited through cell division. *Submitted*, 2016.

[66] N. Dillon, T. Trimborn, J. Strouboulis, P. Fraser, and F. Grosveld. The effect of distance on long-range chromatin interactions. *Molecular Cell*, 1(1):131–139, 1997.

[67] N. Dimitrova, Y. M. Chen, D. L. Spector, and T. de Lange. 53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility. *Nature*, 456(7221):524–528, 2008.

[68] V. Dion and S. M. Gasser. Chromatin movement in the maintenance of genome stability. *Cell*, 152:1355–1364, 2013.

[69] V. Dion, V. Kalck, C. Horigome, B. D. Towbin, and S. M. Gasser. Increased mobility of double-strand breaks requires mec1, rad9 and the homologous recombination machinery. *Nat. Cell Biol.*, 14:502–509, 2012.

[70] V. Dion, V. Kalck, A. Seeber, T. Schleker, and S. M. Gasser. Cohesin and the nucleolus constrain the mobility of spontaneous repair foci. *EMBO reports*, 14(11):984–91, November 2013.

[71] J.R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485:376–80, 2012.

[72] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Oxford: Clarendon Press, 1986.

[73] J. Dostie, T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee, T. A. Honan, E. D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R. D. Green, and J. Dekker. Chromosome conformation capture carbon copy (5c): A massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 16:1299–1309, 2006.

[74] B. Doyle, G. Fudenberg, M. Imakaev, and L. A. Mirny. Chromatin Loops as Allosteric Modulators of Enhancer-Promoter Interactions. *PLoS Computational Biology*, 10(10):e1003867, 2014.

[75] Z. Duan, M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S. Fields, C. A. Blau, and W. S. Noble. A three-dimensional model of the yeast genome. *Nature*, 465:363–367, 2010.

[76] T. M. Dunn, S. Hahn, S. Ogden, and R. F. Schleif. An operator at -280 base pairs that is required for repression of arabad operon promoter: addition of dna helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci.*, 81:5017, 1984.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[77] Behrouz Eslami-Mossallam, Helmut Schiessel, and John van Noort. Nucleosome dynamics: Sequence matters. *Advances in colloid and interface science*, 232:101–113, 2016.

[78] R. Everaers and A. Rosa. Multi-scale modeling of diffusion-controlled reactions in polymers: renormalisation of reactivity parameters. *J. Chem. Phys.*, 136:014902, 2012.

[79] Guillaume J. Filion, Joke G. van Bemmel, Ulrich Braunschweig, Wendy Talhout, Jop Kind, Lucas D. Ward, Wim Brugman, In??s J. de Castro, Ron M. Kerkhoven, Harmen J. Bussemaker, and Bas van Steensel. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell*, 143(2):212–224, 2010.

[80] J. Finch and A. Klug. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.*, 73:1897–1901, 1976.

[81] P. J. Flory. The configuration of real polymer chains. *J. Chem. Phys.*, 17:303, 1949.

[82] B. A. Friedman and C. Yeung. Renormalization group analysis of polymer cyclization with non-equilibrium initial conditions. *Eur. Phys. J., E Soft Matter*, 21:25, 2006.

[83] G. Fudenberg and L. A. Mirny. Higher-order chromatin structure: bridging physics and biology. *Current Opinion in Genetics & Development*, 22(2):115–124, April 2012.

[84] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. Formation of Chromosomal Domains by Loop Extrusion. *bioRxiv*, page 024620, 2015.

[85] D. Thirumalai G. Morrison. Semiflexible chains in confined spaces. *PHYSICAL REVIEW E*, 79:011924, 2009.

[86] M. R. Gartenberg, F. R. Neumann, T. Laroche, M. Blaszczyk, and S. M. Gasser. Sir-mediated repression can occur independently of chromosomal and subnuclear contexts. *Cell*, 119(7):955–67, December 2004.

[87] S. M. Gasser, F. Hediger, a. Taddei, F. R. Neumann, and M. R. Gartenberg. The function of telomere clustering in yeast: The circe effect. *Cold Spring Harbor Symposia on Quantitative Biology*, 69:327–337, 2004.

[88] L. Gehlen, S. M. Gasser, and V. Dion. How broken dna finds its template for repair: a computational approach. *Prog. Theor. Phys. Suppl.*, 191:20–29, 2011.

[89] L. R. Gehlen, G. Gruenert, M. B. Jones, C. D. Rodley, J. Langowski, and J. M. O'Sullivan. Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus*, 3(April 2015):370–383, 2012.

[90] A. Ghosh and N. S. Gov. Dynamics of Active Semiflexible Polymers. *Biophysical Journal*, 107(5):1065–1073, 2014.

[91] Luca Giorgetti, Rafael Galupa, Elphège P. Nora, Tristan Piolot, France Lam, Job Dekker, Guido Tiana, and Edith Heard. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, 157:950–963, 2014.

[92] Anton Goloborodko, John F. Marko, and Leonid Mirny. Mitotic chromosome compaction via active loop extrusion. *bioRxiv*, page 021642, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[93] C. Gong, P. Bongiorno, A. Martins, N. C. Stephanou, H. Zhu, S. Shuman, and M. S. Glickman. Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by ku, ligase d and ligase c. *Nat. Struct. Mol. Biol.*, 12(4):304–312, 2005.

[94] M. Gotta, S. Strahl-Bolsinger, H. Renauld, T. Laroche, B. K. Kennedy, M. Grunstein, and S. M. Gasser. Localization of Sir2p: The nucleolus as a compartment for silent information regulators. *EMBO Journal*, 16(11):3243–3255, 1997.

[95] A. U. Grosberg. *Theoretical and Mathematical Models in Polymer Research*. Academic Press, Boston, 1998.

[96] A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich. The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de physique*, 49:2095–2100, 1988.

[97] A. Y. Grosberg, Y. Rabin, S. Havlin, and A. Neer. Crumpled globule model of the three-dimensional structure of dna. *Europhys. Lett.*, 23:373, 1993.

[98] A. Groth, W. Rocha, A. Verreault, and G. Almouzni. Chromatin challenges during dna replication and repair. *Cell*, 128:721–733, 2007.

[99] T. Guerin, O. Benichou, and Voituriez R. Non-Markovian polymer reaction kinetics. *Nat. Chem.*, 245:568–573, 2013.

[100] T. Guerin, O. Benichou, and Voituriez R. Reactive conformations and non-Markovian cyclization kinetics of a Rouse polymer. *J. Chem. Phys.*, 138:094908, 2013.

[101] T. Guérin, M. Dolgushev, O. Bénichou, R. Voituriez, and A. Blumen. Cyclization kinetics of gaussian semiflexible polymer chains. *Phys. Rev. E*, 90:052601, Nov 2014.

[102] J. E. Haber. Mating-type gene switching in saccharomyces cerevisiae. *Annu. Rev. Genet.*, 32:561–599, 1998.

[103] P. Hahnfeldt, J. E. Hearst, D. J. Brenner, R. K. Sachs, and L. R. Hlatky. Polymer models for interphase chromosomes. *Proc. Natl. Acad. Sci. U.S.A.*, 90:7854, 1993.

[104] H. Hajjoul, S. Kocanova, I. Lassadi, K. Bystricky, and A. Bancaud. Lab-on-chip for fast 3d particle tracking in living cells. *Lab Chip.*, 9:3054–3058, 2009.

[105] H. Hajjoul, J. Mathon, H. Ranchon, I. Goiffon, J. Mozziconacci, B. Albert, P. Carrivain, J.-M. Victor, O. Gadal, Ke. Bystricky, and A. Bancaud. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome research*, 23(11):1829–38, November 2013.

[106] Jonathan D Halverson, Jan Smrek, Kurt Kremer, and Alexander Y Grosberg. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Reports on Progress in Physics*, 77(2):022601, 2014.

[107] L. Harnau, R. G. Winkler, and P. Reineker. Dynamic properties of molecular chains with variable stiffness. *J. Chem. Phys.*, 102:7750, 1995.

[108] M. Hauer, A. Seeber, V. Singh, R. Thierry, J. Amitai, A. Eglinger, Holcman D., T. Owen-Hughes, and Gasser S. Histone degradation in response to dna damage triggers general chromatin decompaction. *Submitted*, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[109] D. W. Heermann. Physical nuclear organization: Loops and entropy. *Current Opinion in Cell Biology*, 23(3):332–337, 2011.

[110] S. Henikoff. Nuclear organization and gene expression: homologous pairing and long-range interactions. *Curr. Opin. Cell Biol.*, 9:388–395, 1997.

[111] P. Heun, T. Laroche, K. Shimada, P. Furrer, and S. M. Gasser. Chromosome dynamics in the yeast interphase nucleus. *Science*, 294(5549):2181–6, 2001.

[112] W. M. Hicks, M. Yamaguchi, and J. E. Haber. Real-time analysis of double-strand dna break repair by homologous recombination. *Proc Natl Acad Sci U S A*, 108:3108–3115, 2011.

[113] Michael Hinczewski and Roland R. Netz. Anisotropic hydrodynamic mean-field theory for semiflexible polymers under tension. *Macromolecules*, 44(17):6972–6985, 2011.

[114] Michael Hinczewski, Xaver Schlagberger, Michael Rubinstein, Oleg Krichevsky, and Roland R. Netz. End-monomer dynamics in semiflexible polymers. *Macromolecules*, 42(3):860–875, 2009.

[115] Tatsuya Hirano and Timothy J. Mitchison. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell*, 79(3):449–458, 1994.

[116] D. Holcman, N. Hoze, and Z. Schuss. Analysis of single particle trajectories: when things go wrong. *arXiv:1502.00286*, 2015.

[117] D. Holcman and I. Kupka. Some questions in computational cellular biology. *Journal of Fixed Point Theory and Applications*, 7:67–83, 2010.

[118] D. Holcman and Z. Schuss. The narrow escape problem. *SIAM Review*, 56:213–257, 2014.

[119] D. Holcman and Z. Schuss. Time scale of diffusion in molecular and cellular biology. *Journal of Physics A: Mathematical and Theoretical*, 47(17):173001, 2014.

[120] C. Horigome, Y. Oma, T. Konishi, R. Schmid, I. Marcomini, M. H. Hauer, V. Dion, M. Harata, and S. M. Gasser. Swr1 and ino80 chromatin remodelers contribute to dna double-strand break perinuclear anchorage site choice. *Molecular Cell*, 55:626–639, 2014.

[121] Chunhui Hou, Li Li, Zhaohui S. Qin, and Victor G. Corces. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Molecular Cell*, 48(3):471–484, 2012.

[122] N. Hozé and D. Holcman. Coagulation-fragmentation for a finite number of particles and application to telomere clustering in the yeast nucleus. *Phys. Lett. A*, 845:5043, 2012.

[123] N. Hoze and D. Holcman. Kinetics of aggregation with a finite number of particles and application to viral capsid assembly. *Journal of Mathematical Biology*, 2014.

[124] N. Hozé, M. Ruault, C. Amoruso, A. Taddei, and D. Holcman. Spatial telomere organization and clustering in yeast saccharomyces cerevisiae nucleus is generated by a random dynamics of aggregation-dissociation. *Mol Biol Cell*, 24:1791–800, 2013.

[125] C. Hyeon and D. Thirumalai. Kinetics of interior loop formation in semiflexible chains. *J. Chem. Phys.*, 124:104905, 2006.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[126] I. Izeddin, V. Récamier, L. Bosanac, I. Cissé, L. Boudarene, C. Dugast-Darzacq, F. Proux, O. Bénichou, R. Voituriez, O. Bensaude, M. Dahan, and X. Darzacq. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife*, 3, 2014.

[127] A. Javer, Z. Long, E. Nugent, M. Grisi, K. Siriwatwetchakul, K. D. Dorfman, P. Cicuta, and M. Cosentino Lagomarsino. Short-time movement of E. coli chromosomal loci depends on coordinate and subcellular localization. *Nature communications*, 4(May):3003, January 2013.

[128] Kuwada NJ Long Z Benza VG Dorfman KD Wiggins PA Cicuta P Lagomarsino MC. Javer A. Persistent super-diffusive motion of escherichia coli chromosomal loci. *Nat Commun.*, 5:3854, 2014.

[129] Siriwatwetchakul K Dorfman KD Cicuta P Cosentino Lagomarsino M. Javer A, Long Z Nugent E Grisi M. Short-time movement of e. coli chromosomal loci depends on coordinate and subcellular localization. *Nat Commun.*, 4:3003, 2013.

[130] J. Jeon and R. Metzler. Fractional brownian motion and motion governed by the fractional langevin equation in confined geometries. *Phys. Rev. E*, 81:021103, 2010.

[131] Q.W. Jin, J. Fuchs, and J. Loidl. Centromere clustering is a major determinant of yeast interphase nuclear organization. *J Cell Sci*, 113 (Pt 11):1903–1912, 2000.

[132] Daniel Jost, Pascal Carrivain, Giacomo Cavalli, and Cédric Vaillant. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic acids research*, 42(15):9553–61, 2014.

[133] Ha BY Jung Y, Jun S. Self-avoiding polymer trapped inside a cylindrical pore: Flory free energy and unexpected dynamics. *Phys Rev E Stat Nonlin Soft Matter Phys.*, 79:011924, 2009.

[134] I. Junier, F. Boccard, and O. Espeli. Polymer modeling of the e. coli genome reveals the involvement of locus positioning and macrodomain structuring for the control of chromosome conformation and segregation. *Proc. Natl Acad. Sci. USA*, 103:12388–12393, 2006.

[135] I. Junier, O. Martin, and F. Képès. Spatial and topological organization of DNA chains induced by gene co-localization. *PLoS Computational Biology*, 6(2), 2010.

[136] M. H. Kagey, J. J. Newman, S. Bilodeau, Y. Zhan, A. David, N. L. Van Berkum, C. C. Ebmeier, J. Goossens, B. Peter, S. S. Levine, D. J Taatjes, J. Dekker, and R. A Young. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314):430–435, 2010.

[137] Y. Kantor and M. Kardar. Anomalous diffusion with absorbing boundary. *Phys Rev. E*, 76:061121, 2007.

[138] E. Kepten, I. Bronshtein, and Y. Garini. Ergodicity convergence test suggests telomere motion obeys fractional dynamics. *Phys. Rev. E*, 83:041919, 2011.

[139] E. Kepten, I. Bronshtein, and Y. Garini. Improved estimation of anomalous diffusion exponents in single-particle tracking experiments. *Phys. Rev. E*, 87:052713, 2013.

[140] R. Kimmich. Strange kinetics, porous media, and NMR. *Chem. Phys.*, 284:253, 2001.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[141] J. Klafter, A. Blumen, and M. F. Shlesinger. Stochastic pathway to anomalous diffusion. *Phys. Rev. A*, 35:3081, 1987.

[142] A. Klemm, H.-P. Müller, and R. Kimmich. NMR microscopy of pore-space backbones in rock, sponge, and sand in comparison with random percolation model objects. *Phys. Rev. E*, 55:4413, 1997.

[143] K. Klenin, H. Merlitz, and J. Langowski. A brownian dynamics program for the simulation of linear and circular dna and other wormlike chain polyelectrolytes. *Biophys J*, 74:780–788, 1998.

[144] S. C. Kou. Stochastic modeling in nanoscale biophysics: Subdiffusion within proteins. *The Annals of Applied Statistics*, 2(2):501–535, June 2008.

[145] S. C. Kou and X. S. Xie. Generalized langevin equation with fractional gaussian noise: Subdiffusion within a single protein molecule. *Phys. Rev. Lett.*, 93:180603, 2004.

[146] F Kouzine, A Gupta, L Baranello, D Wojtowicz, K Ben-Aissa, J Liu, T M Przytycka, and D Levens. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol*, 20(3):396–403, 2013.

[147] Fedor Kouzine, Suzanne Sanford, Zichrini Elisha-Feil, and David Levens. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nature Structural & Molecular Biology*, 15(2):146–154, 2008.

[148] H. Krmer, M. Niemller, M. Amouyal, B. Revet, B. von Wilcken-Bergmann, and B. Mller-Hill. lac repressor forms loops with linear dna carrying two suitably spaced lac operators. *EMBO J.*, 6:1481–1491, 1987.

[149] W. Kuhn. Over the shape of threadlike molecules in solutions. *Kolloid*, 68:2, 1934.

[150] C. Lanctôt, T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature reviews. Genetics*, 8(2):104–15, February 2007.

[151] T. Langenberg, T. Dracz, A. C. Oates, C. Heisenberg, and M. Brand. Analysis and visualization of cell movement in the developing zebrafish brain. *Developmental dynamics : an official publication of the American Association of Anatomists*, 235(4):928–33, April 2006.

[152] J. Langowski. Polymer chain models of dna and chromatin. *Eur. Phys. J. E*, 19:241, 2006.

[153] J. Langowski and D. W. Heermann. Computational modeling of the chromatin fiber. *Semin. Cell Dev. Biol.*, 18(5):659 – 667, 2007.

[154] S. E. Lee, J. K. Moore, A. Holmes, K. Umezu, R.D. Kolodner, and J. E. Haber. Saccharomyces ku70, mre11/rad50 and rpa proteins regulate adaptation to g2/m arrest after dna damage. *Cell*, 94:399–409, 1998.

[155] C. Lemaitre and E. Soutoglou. Double strand break (dsb) repair in heterochromatin and heterochromatin proteins in dsb repair. *DNA Repair (Amst).*, 19:163–8, 2014.

[156] J. E. Lennard-Jones. On the determination of molecular fields. *Proc. R. Soc. Lond. A*, 106:463477, 1924.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[157] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326:289–293, 2009.

[158] M. Lindner, G. Nir, A. Vivante, I. T. Young, and Y. Garini. Dynamic analysis of a diffusing particle in a trapping potential. *Physical Review E*, 87(2):022716, February 2013.

[159] M. Lisby, J. H. Barlow, R. C. Burgess, and R. Rothstein. Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell*, 118(6):699–713, September 2004.

[160] M. Lisby, U. H. Mortensen, and R. Rothstein. Colocalization of multiple dna double-strand breaks at a single rad52 repair centre. *Nat Cell Biol*, 5:572–7, 2003.

[161] M. Lomholt, T. Ambjörnsson, and R. Metzler. Optimal Target Search on a Fast-Folding Polymer Chain with Volume Exchange. *Physical Review Letters*, 95(26):260603, December 2005.

[162] Joseph S. Lucas, Yaojun Zhang, Olga K. Dudko, and Cornelis Murre. 3D trajectories adopted by coding and regulatory DNA elements: First-passage times for genomic interactions. *Cell*, 158(2):339–352, 2014.

[163] W. D. Luedtke and U. Landman. Slip diffusion and Lévy flights of an adsorbed gold nanocluster. *Phys. Rev. Lett.*, 82:3835, 1999.

[164] DaríoG. Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, Hülya Kayserili, JohnM. Opitz, Renata Laxova, Fernando Santos-Simarro, Brigitte Gilbert-Dussardier, Lars Wittler, Marina Borschiwer, StefanA. Haas, Marco Osterwalder, Martin Franke, Bernd Timmermann, Jochen Hecht, Malte Spielmann, Axel Visel, and Stefan Mundlos. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, 161(5):1012–1025, may 2015.

[165] D. M. MacAlpine and G. Almouzni. Chromatin and DNA replication. *Cold Spring Harbor Perspectives in Biology*, 5(8):1–22, 2013.

[166] N. L. Mahy, P. E. Perry, and W. A. Bickmore. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by fish. *J Cell Biol*, 159(5):753–63, 2002.

[167] G. Malherbe and D. Holcman. Search for a dna target site in the nucleus. *Phys. Lett. A.*, 374:466–471, 2010.

[168] G. Malherbe and D. Holcman. Stochastic modeling of gene activation and application to cell regulation. *J Theor Biol.*, 271:51–63, 2010.

[169] J. F. Marko and I. A. Stegun. *The Mitotic Chromosome: Structure and Mechanics, in Genome Organization and Function in the Cell Nucleus.* Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2011.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[170] W. F. Marshall, A. Straight, J. F. Marko, J. Swedlow, A. Dernburg, A. Belmont, A. W. Murray, D. A. Agard, and J. W. Sedat. Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr Biol*, 7(12):930–9, 1997.

[171] C. Massie and I. Mills. Chipping away at gene regulation. *EMBO Rep. Rev.*, 39:337343, 2008.

[172] O. Masui, I. Bonnet, P. Le Baccon, I. Brito, T. Pollex, N. Murphy, P. Hup, E. Barillot, A. S. Belmont, and E. Heard. Live-cell chromosome dynamics and outcome of x chromosome pairing events during es cell differentiation. *Cell*, 145:447–58, 2011.

[173] J. Mateos-Langerak, M. Bohn, W. de Leeuw, O. Giromus, E. M. M. Manders, P. J. Verschure, M. H. G. Indemans, H. J. Gierman, D. W. Heermann, R. van Driel, and S. Goetze. Spatially confined folding of chromatin in the interphase nucleus. *Proc Natl Acad Sci U S A*, 106:3812–7, 2009.

[174] K. Matthews. ladna looping. *Microbiol. Rev.*, 56:123–136, 1992.

[175] Benza VG Bassetti B Dorfman KD Scolari VF Bromek K Cicuta P Lagomarsino MC. Physical descriptions of the bacterial nucleoid at large scales, and their biological implications. *Rep Prog Phys*, 7:75, 2012.

[176] K. J. Meaburn, T. Misteli, and E. Soutoglou. Spatial genome organization in the formation of chromosomal translocations. *Semin. Cancer Biol.*, 17:80, 2007.

[177] P. Meister, L. R. Gehlen, E. Varela, V. Kalck, and S. M. Gasser. Visualizing yeast chromosomes and nuclear architecture. *Methods Enzymol.*, 470:535–567, 2010.

[178] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339(1):1 – 77, 2000.

[179] R. Metzler and J. Klafter. The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *J. Phys. A: Math. Gen.*, 37:R161, 2004.

[180] C. Michaelis, R. Ciosk, and K. Nasmyth. Cohesins: Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids. *Cell*, 91(1):35–45, October 1997.

[181] J. Miné-Hattab and R. Rothstein. Increased chromosome mobility facilitates homology search during recombination. *Nat. Cell Biol.*, 14:510–517, 2012.

[182] J. Miné-Hattab and R. Rothstein. DNA in motion during double-strand break repair. *Trends in cell biology*, 23(11):529–36, November 2013.

[183] L. A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.*, 19:37–51, 2011.

[184] T. Misteli and E. Soutoglou. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nature reviews. Molecular cell biology*, 10(March):243–254, 2009.

[185] Takeshi Mizuguchi, Geoffrey Fudenberg, Sameet Mehta, Jon-Matthew Belton, Nitika Taneja, Hernan Diego Folco, Peter FitzGerald, Job Dekker, Leonid Mirny, Jemima Barrowman, and Shiv I S Grewal. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature*, 516(7531):432–5, 2014.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[186] Florence Mompart, David Robelin, Chantal Delcros, and Martine Yerle-Bouissou. 3D organization of telomeres in porcine neutrophils and analysis of LPS-activation effect. *BMC cell biology*, 14(1):30, 2013.

[187] C. Münkel and J. Langowski. Chromosome structure predicted by a polymer model. *Phys. Rev. E*, 57:5888–5896, 1998.

[188] S. Nagai, K. Dubrana, M. Tsai-Pflugfelder, M. B. Davidson, T. M. Roberts, G. W. Brown, E. Varela, F. Hediger, S. M. Gasser, and N. J. Krogan. Functional targeting of dna damage to a nuclear pore-associated sumo-dependent ubiquitin ligase. *Science*, 322:597–602, 2008.

[189] Takashi Nagano, Yaniv Lubling, Tim J Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D Laue, Amos Tanay, and Peter Fraser. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502:59–64, 2013.

[190] K Nasmyth. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annual review of genetics*, 35:673–745, 2001.

[191] Catherine Naughton, Nicolaos Avlonitis, Samuel Corless, James G Prendergast, Ioulia K Mati, Paul P Eijk, Scott L Cockroft, Mark Bradley, Bauke Ylstra, and Nick Gilbert. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nature structural & molecular biology*, 20(3):387–95, 2013.

[192] Natalia Naumova, Maxim Imakaev, Geoffrey Fudenberg, Ye Zhan, Bryan R Lajoie, Leonid a Mirny, and Job Dekker. Organization of the mitotic chromosome. *Science (New York, N.Y.)*, 342(6161):948–53, 2013.

[193] F. R. Neumann, V. Dion, L. R. Gehlen, M. Tsai-Pflugfelder, R. Schmid, A. Taddei, and S. M. Gasser. Targeted INO80 enhances subnuclear chromatin movement and ectopic homologous recombination. *Genes & development*, 26(4):369–83, February 2012.

[194] M. Nicodemi, B. Panning, and A. Prisco. A Thermodynamic Switch for Chromosome Colocalization. *Genetics*, 179:717721, 2008.

[195] M. Nicodemi and A. Prisco. Thermodynamic Pathways to Genome Spatial Organization in the Cell Nucleus. *Biophysj*, 96(6):2168–2177, 2009.

[196] Mario Nicodemi and Ana Pombo. Models of chromosome structure. *Current opinion in cell biology*, 28:90–95, 2014.

[197] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–5, May 2012.

[198] Anders B Norgaard, Jesper Ferkinghoff-Borg, and Kresten Lindorff-Larsen. Experimental parameterization of an energy function for the simulation of unfolded proteins. *Biophysical journal*, 94(1):182–92, 2008.

[199] E. Nunez, X. Fu, and M. G. Rosenfeld. Nuclear organization in the 3D space of the nucleus - cause or consequence? *Current opinion in genetics & development*, 19(5):424–36, October 2009.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[200] J. M. O'Sullivan, M. D. Hendy, T. Pichugina, G. C. Wake, and J. Langowski. The statistical-mechanics of chromosome conformation capture. *Nucleus (Austin, Tex.)*, 4(5):390–8, 2013.

[201] P. Oza, S. L. Jaspersen, A. Miele, J. Dekker, and C. L. Peterson. Mechanisms that regulate localization of a dna double strand break to the nuclear periphery. *Genes Dev*, 23:912–927, 2009.

[202] R. W. Pastor, R. Zwanzig, and A. Szabo. Diffusion limited first contact of the ends of a polymer: Comparison of theory with simulation. *J. Chem. Phys.*, 105:3878–3882, 1996.

[203] C. L. Peterson and G. Almouzni. Nucleosome dynamics as modular systems that integrate DNA damage and repair. *Cold Spring Harbor perspectives in biology*, 5(9), September 2013.

[204] T. D. Petes. Yeast ribosomal dna genes are located on chromosome xii. *Proceedings of the National Academy of Sciences*, 76(1):410–414, 1979.

[205] E. P. Petrov, T. Ohrt, R. G. Winkler, and P. Schwille. Diffusion and segmental dynamics of double-stranded DNA. *Physical Review Letters*, 97(25):1–4, 2006.

[206] A. Podtelezhnikov and A. Vologodskii. Simulations of polymer cyclization by brownian dynamics. *Macromolecules*, 30:6668, 1997.

[207] Tim Pollex and Edith Heard. Recent advances in X-chromosome inactivation research. *Current opinion in cell biology*, 24(6):825–32, dec 2012.

[208] A. F. Price and A. D. D'Andrea. Chromatin remodeling at dna double-strand breaks. *Cell*, 152:1344–54, 2013.

[209] M. Ptashne. Gene regulation by proteins acting nearby and at a distance. *Nature*, 322:697, 1986.

[210] W. K. Purves, D. E. Sadava, G. H. Orians, and H. C. Heller. *Life: The Science of Biology Seventh Edition 7th Edition*. Sinauer Associates and W. H. Freeman, 2004.

[211] H. Qian, M. P. Sheetz, and E. L. Elson. Single particle tracking. analysis of diffusion and flow in two-dimensional systems. *Biophys. J.*, 60:91021, 1991.

[212] Angelo R. and Christophe Z. Chapter nine - computational models of large-scale genome architecture. In *New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals*, volume 307 of *International Review of Cell and Molecular Biology*, pages 275 – 349. Academic Press, 2014.

[213] C. Rabl. ber zellteilung. *Morphol. Jahrbuch.*, 10:214330, 1885.

[214] SuhasS.P. Rao, MiriamH. Huntley, NevaC. Durand, ElenaK. Stamenova, IvanD. Bochkov, JamesT. Robinson, AdrianL. Sanborn, Ido Machol, ArinaD. Omer, EricS. Lander, and ErezLieberman Aiden. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 159(7):1665–1680, dec 2014.

[215] J. Reingruber, E. Abad, and D. Holcman. Narrow escape time to a structured target located on the boundary of a microdomain. *J. Chem. Phys.*, 130:094909, 2009.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[216] L. F. Richardson. Atmospheric diffusion shown on a distance-neighbour graph. *Proc. Roy. Soc.*, 110:709, 1926.

[217] D. R. Rines, D. Thomann, J. F. Dorn, P. Goodwin, and P. K. Sorger. Live cell imaging of yeast. *Cold Spring Harbor protocols*, 30:1026–1041, 2011.

[218] L. Ringrose, S. Chabanis, O. P. Angrand, C. Woodroofe, and A. F. Stewart. Quantitative comparison of dna looping in vitro and in vivo: chromatin increases effective dna flexibility at short distances. *EMBO J.*, 18:6630, 1999.

[219] H. Risken. *The Fokker-Planck Equation.* Springer-Verlag New York, LLC, 1996.

[220] C. C. Robinett, A. Straight, G. Li, C. Willhelm, G. Sudlow, A. Murray, and A. S. Belmont. In vivo localization of dna sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol*, 135:1685–700, 1996.

[221] C. D. M. Rodley, F. Bertels, B. Jones, and J. M. O'Sullivan. Global identification of yeast chromosome interactions using genome conformation capture. *Fungal Genet Biol*, 46:879–86, 2009.

[222] A. Rosa, N. B. Becker, and R. Everaers. Looping probabilities in model interphase chromosomes. *Biophys J*, 98(11):2410–9, 2010.

[223] A. Rosa and R. Everaers. Structure and dynamics of interphase chromosomes. *PLOS Comput Biol*, 4:e1000153, 2008.

[224] P. E. Rouse. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J. Chem. Phys.*, 21:1272, 1953.

[225] M. Ruault, A. De Meyer, I. Loiodice, and A. Taddei. Clustering heterochromatin: Sir3 promotes telomere clustering independently of silencing in yeast. *J Cell Biol*, 192(3):417–31, 2011.

[226] Michael Rubinstein and R Colby. *Polymers physics*, volume 767. Oxford Oxford, UK, 2003.

[227] Michael J Rust, Mark Bates, and Xiaowei W Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods*, 3(10):793–795, 2006.

[228] P. G. Saffman and M. Delbrck. Brownian motion in biological membranes. *Proceedings of the National Academy of Sciences*, 72(8):3111–3113, 1975.

[229] T. Sakaue and E. Raphael. Polymer chains in confined spaces and flow-injection problems: Some remarks. *Macromolecules*, 39:2621–2628, 2006.

[230] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman and Hall, New York, 1994.

[231] Adrian L. Sanborn, Suhas S. P. Rao, Su-Chen Huang, Neva C. Durand, Miriam H. Huntley, Andrew I. Jewett, Ivan D. Bochkov, Dharmaraj Chinnappan, Ashok Cutkosky, Jian Li, Kristopher P. Geeting, Andreas Gnirke, Alexandre Melnikov, Doug McKenna, Elena K. Stamenova, Eric S. Lander, and Erez Lieberman Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, 112(47):201518552, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[232] I. F. Sbalzarini and P. Koumoutsakos. Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.*, 151:182–195, 2005.

[233] H. Scher and E. W. Montroll. Anomalous transit-time dispersion in amorphous solids. *Phys. Rev. B*, 12:2455, 1975.

[234] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. Nih image to imagej: 25 years of image analysis. *Nat. Methods*, 9:671–675, 2012.

[235] H. Schober, V. Kalck, M. A. Vega-Palas, G. Van Houwe, D. Sage, M. Unser, M. R. Gartenberg, and S. M. Gasser. Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Res*, 18:261–271, 2008.

[236] Z. Schuss. *Theory and Applications of Stochastic Differential Equations.* Wiley Series in Probability and Statistics - Applied Probability and Statistics Section, 1980.

[237] Z. Schuss. *Diffusion and Stochastic Processes. An Analytical Approach.* Springer-Verlag, New York, NY, 2009.

[238] Z. Schuss. *Brownian Dynamics at Boundaries and Interfaces In Physics, Chemistry, and Biology Applied Mathematical Sciences.* Springer-Verlag, New York, NY, 2013.

[239] Z. Schuss, A. Singer, and D. Holcman. The narrow escape problem for diffusion in cellular microdomains. *Proc Natl Acad Sci U S A*, 104:16098–16103, 2007.

[240] Wibke Schwarzer and Fran??ois Spitz. The architecture of gene expression: Integrating dispersed cis-regulatory modules into coherent regulatory domains. *Current Opinion in Genetics and Development*, 27:74–82, 2014.

[241] Vittore F. Scolari and Marco Cosentino Lagomarsino. Combined collapse by bridging and self-adhesion in a prototypical polymer model inspired by the bacterial nucleoid. *Soft Matter*, 11:1677–1687, 2015.

[242] Andrew Seeber, Vincent Dion, and Susan M Gasser. Checkpoint kinases and the INO80 nucleosome remodeling complex enhance global chromatin mobility in response to DNA damage. *Genes & development*, 27(18):1999–2008, September 2013.

[243] P. Sengupta, S. Van Engelenburg, and J. Lippincott-Schwartz. Visualizing cell structure and function with point-localization superresolution imaging. *Developmental cell*, 23(6):1092–102, December 2012.

[244] T. Sexton, F. Bantignies, and G. Cavalli. Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation. *Seminars in cell & developmental biology*, 20(7):849–55, September 2009.

[245] Tom Sexton, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3):458–72, feb 2012.

[246] O. Shukron and D. Holcman. Tads structure and gene regulation embedded in hi-c data revealed by polymer reconstruction and stochastic simulations. *Submitted*, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[247] R. Shusterman, S. Alon, T. Gavrinyov, and O. Krichevsky. Monomer dynamics in double- and single-stranded dna polymers. *Phys. Rev. Lett.*, 92(4):048303, 2004.

[248] A. Singer and Z. Schuss. Activation through a narrow opening. *SIAM J. Appl. Math.*, 68:98, 2007.

[249] A. Singer, Z. Schuss, A. Osipov, and D. Holcman. Partially reflected diffusion. *SIAM J. Appl. Math.*, 68:844–868, 2008.

[250] M. Slutsky, M. Kardar, and L. A. Mirny. Diffusion in correlated random potentials, with applications to DNA. *Phys. Rev. E*, 69:061903, 2004.

[251] M. Slutsky and L. A. Mirny. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophysical journal*, 87(December):4021–4035, 2004.

[252] G. Smeenk and H. van Attikum. The chromatin response to DNA breaks: leaving a mark on genome integrity. *Annual review of biochemistry*, 82:55–80, 2013.

[253] Sevil Sofueva, Eitan Yaffe, Wen-Ching Chan, Dimitra Georgopoulou, Matteo Vietri Rudan, Hegias Mira-Bontenbal, Steven M Pollard, Gary P Schroth, Amos Tanay, and Suzana Hadjur. Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO journal*, 32(24):3119–29, 2013.

[254] I. M. Sokolov. Cyclization of a polymer: First-passage problem for a non-markovian process. *Macromolecules*, 90:080601, 2003.

[255] E. Soutoglou and T. Misteli. Mobility and immobility of chromatin in transcription and genome stability. *Annu. Rev. Cell Dev. Biol.*, 17:435–442, 2007.

[256] J. Stampe and I. M. Sokolov. Cyclization of a polymer with charged reactive end groups. *J. Chem. Phys.*, 114:5043, 2001.

[257] M. Stevens. Simple simulations of dna condensation. *Biophys J.*, 80:130, 2001.

[258] Jonathan Strecker, Gagan D. Gupta, Wei Zhang, Mikhail Bashkurov, Marie-Claude Landry, Laurence Pelletier, and Daniel Durocher. DNA damage signalling targets the kinetochore to promote chromatin mobility. *Nature Cell Biology*, 18(3):281–90, 2016.

[259] H. Sutherland and W. A. Bickmore. Transcription factories: gene expression in unions? *Nat Rev Genet*, 10(7):457–66, 2009.

[260] Orsolya Symmons, Veli Vural Uslu, Taro Tsujimura, Sandra Ruf, Sonya Nassari, Wibke Schwarzer, Laurence Ettwiller, and Fran??ois Spitz. Functional and topological characteristics of mammalian regulatory domains. *Genome Research*, 24(3):390–400, 2014.

[261] A. Szabo, K. Schulten, and Z. Schulten. First passage time approach to diffusion controlled reactions. *J. Chem. Phys.*, 72:4350–4357, 1980.

[262] J. W. Szostak, T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl. The double-strand-break repair model for recombination. *Cell*, 33(1):25 – 35, 1983.

[263] A. Taddei and S. M. Gasser. Structure and function in the budding yeast nucleus. *Genetics*, 192:107–29, 2012.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[264] A. Taddei, F. Hediger, F. R. Neumann, C. Bauer, and S. M. Gasser. Separation of silencing from perinuclear anchoring functions in yeast ku80, sir4 and esc1 proteins. *EMBO J*, 23(6):1301–12, 2004.

[265] A. Taddei, H. Schober, and S. M. Gasser. The budding yeast nucleus. *Cold Spring Harbor perspectives in biology*, 2:a000612, 2010.

[266] A. Taddei, G. Van Houwe, F. Hediger, V. Kalck, F. Cubizolles, H. Schober, and S. M. Gasser. Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature*, 441:774–778, 2006.

[267] S. M. Tan-Wong, H. D. Wijayatilake, and N. J. Proudfoot. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes & development*, 23(22):2610–24, November 2009.

[268] M. Tark-Dame, H. Jerabek, E. M. M. Manders, D. W. Heermann, and R. van Driel. Depletion of the Chromatin Looping Proteins CTCF and Cohesin Causes Chromatin Compaction: Insight into Chromatin Folding by Polymer Modelling. *PLoS Computational Biology*, 10(10):e1003877, 2014.

[269] P. Therizols, T. Duong, B. Dujon, C. Zimmer, and E. Fabre. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc Natl Acad Sci U S A*, 107:2025–2030, 2010.

[270] Guido Tiana, Assaf Amitai, Tim Pollex, Tristan Piolot, David Holcman, Edith Heard, and Luca Giorgetti. Structural Fluctuations of the Chromatin Fiber within Topologically Associating Domains. *Biophysical Journal*, 110(6):1234–1245, 2016.

[271] C. Ting, J. Ding, and J. Z. Y. Chen. Mean first-passage times of looping of polymers with intrachain reactive monomers: Lattice monte carlo simulations. *Macromolecules*, 39:5540–5545, 2006.

[272] H. Tjong, K. Gong, and F. Alber. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.*, 10:1295–1305, 2012.

[273] N. M. Toan, G. Morrison, C. Hyeon, and D. Thirumalai. Cyclization of rouse chains at long- and short-time scales. *J. Chem. Phys. B.*, 112:6094–6106, 2008.

[274] N. Tokuda, T. P. Terada, and M. Sasai. Dynamical modeling of three-dimensional genome organization in interphase budding yeast. *Biophys J*, 102:296–304, 2012.

[275] I. M. Toli ć Nørrelykke, E. L. Munteanu, G. Thon, L. Oddershede, and K. Berg-Sørensen. Anomalous diffusion in living yeast cells. *Phys. Rev. Lett.*, 93:078102, 2004.

[276] J. Torres-Rosell, I. Sunjevaric, G. De Piccoli, M. Sacher, N. Eckert-Boulet, R. Reid, S. Jentsch, R. Rothstein, L. Aragón, and M. Lisby. The Smc5-Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. *Nature cell biology*, 9(8):923–31, August 2007.

[277] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36:823841, 1930.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[278] S. Vakeroudis, M. Yor, and D. Holcman. The mean first rotation time of a planar polymer. *J. Stat. Phys.*, 143:1074–1095, 2011.

[279] J. S. Verdaasdonk, P. A. Vasquez, R. M. Barry, T. Barry, S. Goodwin, M. G. Forest, and K. Bloom. Centromere tethering confines chromosome domains. *Molecular cell*, 52(6):819–31, December 2013.

[280] Matteo Vietri Rudan, Christopher Barrington, Stephen Henderson, Christina Ernst, Duncan T. Odom, Amos Tanay, and Suzana Hadjur. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Reports*, 10(8):1297–1309, 2015.

[281] E V Volpi, E Chevret, T Jones, R Vatcheva, J Williamson, S Beck, R D Campbell, M Goldsworthy, S H Powis, J Ragoussis, J Trowsdale, and D Sheer. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *Journal of cell science*, 113 (Pt 9:1565–1576, 2000.

[282] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acids. *Nature*, 171(4356):737–738, 1953.

[283] S. C. Weber, A. J. Spakowitz, and J. A. Theriot. Nonthermal ATP-dependent fluctuations contribute to the in vivo motion of chromosomal loci. *Proceedings of the National Academy of Sciences of the United States of America*, 109(19):7338–43, May 2012.

[284] S. C. Weber, J. A. Theriot, and A. J. Spakowitz. Subdiffusive motion of a polymer composed of subdiffusive monomers. *Phys. Rev. E*, 82:011913, 2010.

[285] S.C. Weber, M.A. Thompson, W. Moerner, A.J. Spakowitz, and J.A. Theriot. Analytical tools to distinguish the effects of localization error, confinement, and medium elasticity on the velocity autocorrelation function. *Biophys. J.*, 102:2443–2450, 2012.

[286] George H Weiss, John T Bendler, and Michael F Shlesinger. Polymer melt dynamics model with a relaxation time exponent of 10/3. *Macromolecules*, 21(2):521–523, 1988.

[287] G. Wilemski and M. Fixman. Diffusion-controlled intrachain reactions of polymers. i theory. *J. Chem. Phys.*, 60:866–877, 1974.

[288] G. Wilemski and M. Fixman. Diffusion-controlled intrachain reactions of polymers. ii results for a pair of terminal reactive groups. *J. Chem. Phys.*, 60:878–890, 1974.

[289] J. H. Wilson, W. Y. Leung, G. Bosco, D. Dieu, and J. E. Haber. The frequency of gene targeting in yeast depends on the number of target copies. *Proceedings of the National Academy of Sciences of the United States of America*, 91(January):177–181, 1994.

[290] H. Wong, J.-M. Arbona, and C. Zimmer. How to build a yeast nucleus. *Nucleus (Austin, Tex.)*, 4(5):361–6, 2013.

[291] H. Wong, H. Marie-Nelly, S. Herbert, P. Carrivain, H. Blanc, R. Koszul, E. Fabre, and C. Zimmer. A predictive computational model of the dynamic 3d interphase yeast nucleus. *Current Biology*, 22:1881–1890, 2012.

[292] A. J. Wood, A. F. Severson, and B. J. Meyer. Condensin and cohesin complexity: the expanding repertoire of functions. *Nat Rev Genet*, 11:391–404, 2010.

bioRxiv preprint doi: https://doi.org/10.1101/076661; this version posted January 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[293] N. Xu, S. Donohoe, M. ans Silva, and J. Lee. Evidence that homologous x-chromosome pairing requires transcription of ctcf protein. *Nat. Genet.*, 39:13901396, 2007.

[294] M. Yor. A propos de l'inverse du mouvement brownian dans rn $n \geq 3$. *Annales de l'I.H.P. section B*, 21:27–38, 1985.

[295] A. Zidovska, D. A. Weitz, and T. J. Mitchison. Micron-scale coherence in interphase chromatin dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39):15555–60, 2013.

[296] B. H. Zimm. Dynamics of polymer molecules in dilute solution: Viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.*, 24:269, 1956.

[297] C. Zimmer and E. Fabre. Principles of chromosomal organization: lessons from yeast. *J. Cell Biol.*, 192(5):723–733, 2011.

[298] Y. Ziv, D. Bielopolski, Y. Galanty, C. Lukas, Y. Taya, D. C. Schultz, J. Lukas, S. Bekker-Jensen, J. Bartek, and Y. Shiloh. Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nature cell biology*, 8(8):870–6, August 2006.

[299] R. Zwanzig. Diffusion-controlled ligand binding to spheres partially covered by receptors: an effective medium treatment. *Proceedings of the National Academy of Sciences of the United States of America*, 87(15):5856–5857, 1990.