
# Università degli Studi di Napoli "Federico II"

## Dipartimento di Scienze Fisiche


*Doctorate School in Fundamental and Applied Physics*
*XXIII cycle*

Anno Accademico 2009/2010

**Ph.D. Thesis**

# Statistical Mechanics of Genome Regulation: the case of X Chromosome Inactivation


Relatore:

**Prof. Mario Nicodemi**

Candidato:

**Antonio Scialdone**

*A Sandro*

"La speranza è una trappola, è una brutta parola, non si deve usare, è una trappola inventata dai padroni che ti dicono state zitti, state boni, pregate, ché avrete la vostra ricompensa nell'aldilà, andate a casa, abbiate speranza. Mai avere speranza, la speranza è una trappola, è una cosa infame!"

*Mario Monicelli*

# Contents

# Introduction

Female mammal cells have two copies of X chromosome, one inherited from the father, one from the mother. Yet, to avoid a lethal overproduction of X linked products, only one X chromosome must be active. For this reason, in all female mammal cells, one of the two X chromosomes is inactivated whereas the other remains active; and the choice of the inactive X is completely random. The cellular process leading to this result is called X Chromosome Inactivation (XCI). It is still unknown how the cell can carry out and control such a chromosome-wide stochastic regulatory mechanism, that entails a series of complex events, including the spatial regulation of X chromosomes, and the "counting&choice", which makes the cell "count" the number of X's and randomly "choose" the X to inactivate. The interest of such issues goes well beyond XCI, since it is only the best studied case of random monoallelic expressed genes. This is a class including about 10% of our genes, where out of two alleles one is randomly selected and inactivated, with important and poorly understood examples, ranging from our immune system to the olfactory apparatus. For this reason, XCI has recently attracted substantial interests in the scientific community, with a number of publications in top science journals.

The aim of my PhD research project was to discover the mechanisms behind XCI, by means of theoretical models from Statistical Mechanics based on experimental data. The first part of the project was focused on the mechanisms underlying the X chromosome spatial organization during XCI process. For instance, at the beginning of XCI, X chromosomes have been shown to

recognize each other among tens of different chromosomes, and to get colocalized. The molecular bases of this event are currently unknown. In chapter 2 we propose a DNA "passive shuttling" mechanism which can explain this phenomenon. It works by means of random contacts between a DNA locus and its target (another DNA locus or a nuclear structure) which can be made stable by diffusing molecular mediators able to bridge them. Such a mechanism is reliable and leads to a thermodynamically stable colocalization only if the concentration/affinity of the molecules overcome precise thresholds, otherwise the DNA and its target independently diffuse.

In the second part we have investigated the "counting&choice" mechanism (chapter 3): how is the cell able to realize that two X's are present? how does the random, mutually exclusive choice of the inactive X occur? On the basis of very recent experiments, we postulated the action of two types of molecules which interact with precise DNA sequences and have a mutual affinity. Specific thresholds exist for the molecule mutual affinity and concentration above which the molecules self-assemble into two clusters, one per type, each bound to a different X. So, the symmetry between the X's is broken: one is marked for the inactivation, the other to stay active. This model can well reproduce also the complex 3D spatial architecture of the two X's, which was experimentally analyzed very recently [85].

For the first time, quantitative models were proposed able to describe in a unified framework the early stages of XCI and to reproduce the experimental scenario so far emerged (chapter 4). All the mechanisms we propose can be reliably controlled by the cell by tuning some parameters (such as the amount of specific types of molecules, or their DNA binding affinity) which are known to underlie the basic genome regulatory strategies.

By means of analytic/computational techniques we investigate in depth these models and pointed out a large set of very interesting and non-trivial predictions (for instance, the non-linear effect of DNA deletions in X chromosome pairing, the key role of specific DNA regions for "counting&choice", the pairing configuration of X chromosomes in multiploid cells, etc.). All

these predictions could be useful to design new targeted experiments, which may help clarifying the still mysterious aspects of XCI. Important experimental confirmations of the models have already been found: e.g., possible candidates for the envisaged molecular binders have been recently discovered [12, 24, 25, 92] and the existence of clusters of molecular binding sites in specific DNA regions was confirmed [25, 24].

The models use very basic, minimal molecular ingredients, without the need for specifying the details of the intervening biochemical interactions, as they are rooted in general thermodynamic mechanisms. Thus, it is suggested that these models may describe general regulatory mechanisms the cell can use to manage the genome functioning, and that their validity can go beyond XCI to be applied to other cellular processes that involve similar phenomena (e.g., random monoallelic expression, meiotic chromosome pairing, etc.).

A close interplay between theory and experiments, was guaranteed in our project by the collaboration with the experimental group of Prof. J.T. Lee from Harvard Medical School, USA.

This thesis reports on the results we have achieved during this research project and is based on the following papers:

- A. Scialdone and M. Nicodemi, *Plos Computational Biology* 4:e1000244 (2008)
- A. Scialdone and M. Nicodemi, *Journal of Biomed. and Biotech.* 2009:516723 (2009)
- A. Scialdone and M. Nicodemi, *Advances in Complex Systems* 13:367 (2010)
- A. Scialdone and M. Nicodemi, *Development* 137:3877 (2010)
- A. Scialdone and M. Nicodemi, *EuroPhysics Letters* 92:20002 (2010)
- V. Bianco, A. Scialdone and M. Nicodemi, *in preparation* (2010)
- A. Scialdone, I. Cataudella, A. Prisco and M. Nicodemi, *in preparation* (2010)

# Chapter 1

# The X-Chromosome Inactivation

In this first chapter, we will briefly discuss the biology of X Chromosome Inactivation (XCI), by reporting on the recent discoveries on the molecular bases of it.

## 1.1 Counting&Choosing the X's

The sex of most mammals is decided by a single couple of chromosomes, the X and Y. Males are characterized by the presence of a X and a Y chromosome, while females have two X's. The variation in the number of X chromosomes poses a problem: if X-linked genes were expressed equally well in each sex, females would have twice as much of each product as males. Though, this situation would be lethal for female cells, which, thus, have evolved dosage compensation mechanisms to equalize the dosage of X-chromosome gene products to that of males.

Female mammals achieve dosage compensation by blocking the transcription of one of their two X chromosomes, by a process known as *X Chromosome Inactivation* (XCI). In fact, during the very early stage of development of a female embryo, when it consists of a few thousand cells, one of the two X

chromosomes in each cell becomes highly condensed. The condensed X chromosome can be easily seen under the light microscope in interphase cells: it was originally called a "Barr body". This very dense structure hinders the transcription of the X chromosome, which, thus, will be silenced. The result is that females will have one active X chromosome, which is the same situation found in males.

In all mammals, except for marsupials [2], the X to inactivate is randomly chosen among the maternally inherited one (Xm) or the paternally inherited one (Xp). Once either Xm or Xp has been inactivated, it remains silent throughout all subsequent cell divisions of that cell and its progeny, indicating that the inactive state is faithfully maintained through many cycles of DNA replication and mitosis [1]. Because X-Inactivation is random and takes place after several thousand cells have already formed in the embryo, every female is a mosaic of groups of cells, in which either Xm or Xp is silenced. Since these groups of sister cells tend to remain close together during the later stages of development, they are distributed in small clusters in the adult organism. For instance, this is evident in the red and black "tortoise-shell" coat coloration of some female cats (fig. 1.1). In these cats, one X chromosome carries a gene that produces red hair color, and the other X chromosome carries an allele of the same gene that results in black hair color. The random X-inactivation produces patches of cells of two distinctive colors. In contrast to the females, male cats of this genetic stock are either solid red or solid black, depending on which allele of the hair color gene they have on their only X chromosome.

Summarizing, XCI is a process whereby the cell "counts" the X chromosome it has. And, whenever it realizes that two X's are present, it is able to "choose" just one of them, in a random way, and to inactivate it.

What are the physical mechanisms underlying these amazing abilities of the cell? This open issue, will be one of the key question we will try to answer

[1] Cellular division leading to the formation of two daughter cells from one mother cell. The mother as well as the daughter cells have the same amount of DNA.

Figure 1.1: Red and black "tortoise-shell" coat coloration of some female cats is an effect of X Chromosome Inactivation.

in the following chapters.
In recent years XCI attracted lots of interests in the scientific community, both because it involves regulatory mechanisms which intervene in many other cellular processes (see below), and because of its clinical importance, as misfunctions in it can result in serious diseases [1].

## 1.2 A few Molecular Biology details

Let us give some more details on the DNA sequences and on the molecules which are known to be involved in XCI.
XCI is initiated and spreads from a single site in the middle of X chromosome, the so called *X Inactivation Center* (*Xic*) (fig. 1.2). In particular, this region contains the gene *Xist*, which codes for a RNA molecule fundamental for XCI. The *Xist* RNA is not translated into a protein and remains in the nucleus, where it eventually coats the entire inactive X chromosome. In fact, the transcription status of *Xist* gene, is strictly correlated with the active/silenced status of the X chromosome carrying it: before the inactivation process starts, the *Xist*'s on both genes are in a "poised" state, and are characterized by a low transcription level; yet, as soon as XCI is triggered,

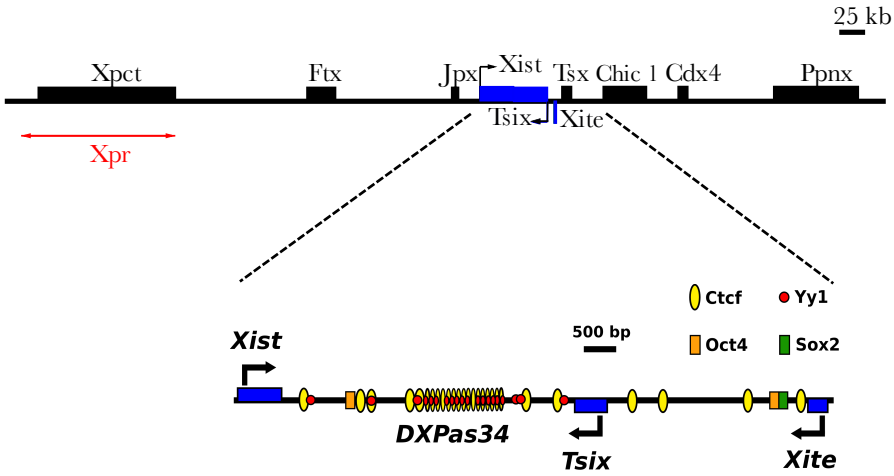


Figure 1.2: Map of *X Inactivation Center*. The regions involved in X chromosome pairing, say *Tsix/Xite* genes and *Xpr*, are highlighted. In the enlargement of the *Tsix/Xite* region the discovered binding sites for Ctcf, Oct4, Yy1 and Sox2 proteins are marked. Ctcf and Oct4 have been proved to be necessary for the pairing of *Tsix/Xite* regions.

while *Xist* from the active X is completely silenced, *Xist* from the inactive X becomes highly transcribed.

*Xic* also includes some other genes (e.g., *Tsix*. *Xite*, etc., see fig. 1.2) whose role in XCI is not totally clear [4, 67]. We will talk about some of them later on.

Clusters of binding sites for several regulatory molecules are also found within *Xic*, and they have a primary role in XCI. For instance, Ctcf and Oct4 proteins bind several DNA sites nearby the *Tsix* and *Xite* genes (see fig. 1.2), and many experiments show that their presence is necessary for a proper XCI to occur [12, 92, 24].

XCI consists of several stages, during which the X chromosomes, and, in particular their *Xic*'s regions, are involved in a complex series of regulatory mechanisms. In the next section we will focus on one of these, occurring at the early stages of XCI: the "pairing" of the two *Xic*'s.

## 1.3 X chromosomes pair off

A crucial initial step that occurs during XCI is the physical colocalization of the two X chromosomes [93, 5, 3] (see fig. 1.3): during the early days of cell differentiation, at the very beginning of XCI, the X chromosomes recognize each other and pair, among tens of other different chromosomes. In particular, their *Xic* regions are in close contact. This pairing event is a preliminary step for XCI and is strictly needed for it; in fact, disruptions of pairing induce XCI failure and cell death [93, 5, 3, 24].

Experiments have pointed out two major regions of colocalization within the *Xic*: a sequence between *Tsix* and *Xite* genes close to *Xist* [93, 5] (see fig. 1.2); and a segment located several hundred of kilobases upstream, named *Xpr* [3] (see fig. 1.2), whose colocalization is independent from *Tsix/Xite* and occurs at an earlier stage. While the specific role of these regions is still under investigation, several details of *Tsix/Xite* have been elucidated.

Its colocalization requires some few kilobase long DNA subfragments and two known Zinc-finger proteins[2], Ctcf and Oct4, having several DNA binding domains which can bind those subfragments at multiple and clustered sites [93, 92, 24] (see fig. 1.2). Interestingly, removal of these binding sites as well as knock-down of these two proteins, severely affect pairing, showing that the binding of these proteins is strictly needed [93, 92, 24].

Moreover, it has been discovered that the inhibition of transcription of *Tsix* and *Xite* disrupts the formation of X-X couples. On these experimental bases, it has been proposed that the X chromosome interaction is mediated by a transiently stable "RNA-protein bridge" at these specific *Xic* sites [92]. Importantly, the insertion on a different chromosome of the mentioned *Tsix/Xite* and *Xpr* segments of *Xic* induces the pairing of X with that chromosome [93, 3, 92]. A still unexplained result is that deletion/insertions, including pairing regions, affect the strength of pairing according to their length, i.e.,

[2] Zinc-Finger proteins are a class of proteins which can recognize and bind to specific DNA sequences thanks to a particular structure (the "Zinc-Finger") including one or more zinc atoms.

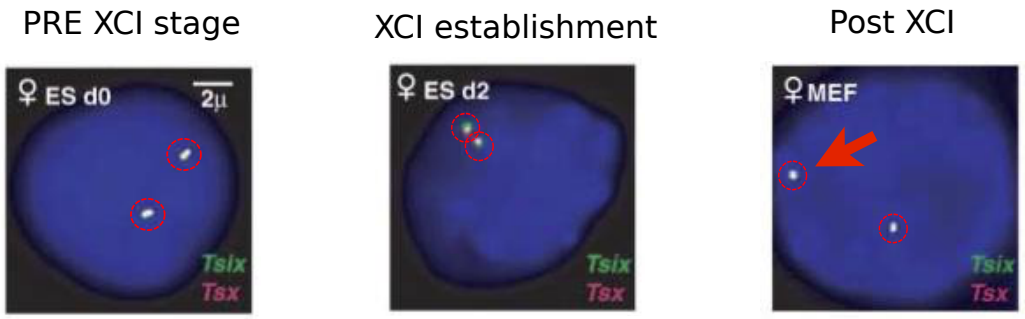


Figure 1.3: Images of real cell nuclei (in blue) at different stages of X Chromosome Inactivation (XCI). The two spots (circled in red) are *Tsix/Xite* regions from the two X chromosomes. As soon as XCI starts, *Tsix/Xite* loci get paired (central panel), while after XCI has occurred, they break up (right panel). In particular, at the end of XCI, the designated active X is located near the nuclear membrane (red arrow in right panel, see section 1.5). Images adapted from [93].

longer deletions exhibit weaker pairing [93]. Consistently, longer insertions of *Xic* pairing segments on an autosome (non sex chromosome) produce stronger X-autosome pairing; and, in females, X-autosome interaction competes with X-X pairing [93].

As deletions of those DNA regions and mutations of Ctcf/Oct4 disrupt colocalization, these elements are thought to be necessary components of the "pairing machinery". Yet, are these ingredients sufficient? Which are the mechanical bases and the physical requirements producing X chromosome recognition and colocalization? *Xic* pairing must occur in a precise time window: how can a cell orchestrate the timing of this process? And how is X chromosome pairing related to the "counting&choice" of X chromosomes? All these crucial questions are still unanswered.

## 1.4 XCI triggering and *Xic* 3D spatial configuration

As we have seen, female cell is somehow able to "count" the X chromosomes and to "choose" randomly the one to inactivate. A number of experiments have been carried out to clarify which are the "controlling factors" for "counting&choice" process, that eventually results in the upregulation of *Xist* gene on the future inactive X and marks the initiation of XCI process.

These experiments assessed the effects on "counting&choice" of heterozygous as well as homozygous [3] deletions/insertions of DNA sequence within the *Xic* (see fig.1.4). In particular, we summarize here the results of four deletions, namely $\Delta 65kb$ [15], $Tsix^{\Delta CpG}$ [47], $Xite^{\Delta L}$ [65] and $\Delta Jpx$ [84], whose location and length is illustrated in fig. 1.4.

The $\Delta 65kb$ deletion removes 65 kilobases of DNA in the *Xic* region relevant to the chromosome activation [15]. $\Delta 65kb$ causes nonrandom inactivation of the deleted X in heterozygous XX cells [15], and determines the inactivation of the unique X chromosome in male XY cells [58]: so, interestingly, the X chromosome bearing the deletion is not active, not even in male cells.

The outcomes in male cells is, however, drastically different in the case of shorter deletions. Importantly, shorter deletions nested into the $\Delta 65kb$ have been described that cause X inactivation in male cells. The $\Delta AS$ deletion [86] is accompanied by a minimal, but detectable, level of ectopic X inactivation. A 1.2 kb deletion called $\Delta 34$ [86], spanning an array of binding sites for Ctcf protein also implicated in the X random choice (see also [25, 16]), causes XCI to initiate in a significant proportion of male cells [86]. Finally, the $\Delta AV$ deletion [86], determines ectopic X inactivation in male cells with high efficiency.

The analysis of two smaller nonoverlapping deletions within the above mentioned $\Delta 65kb$ sequence, namely, the $Tsix^{\Delta CpG}$ deletion and the $Xite^{\Delta L}$,

[3] Heterozygous deletions involve only one of two homologous chromosomes (the X chromosomes in the present case), homozygous deletions both of them.

added further information. While in heterozygous XX cells the $Tsix^{\Delta CpG}$ deletion causes nonrandom inactivation of the deleted X, in male XY cells the $Tsix^{\Delta CpG}$ deleted X remains active [47]. Analogous results are obtained with the $Xite^{\Delta L}$ deletion [65]. It is very interesting the outcome of the homozygous $Tsix^{\Delta CpG}$ female XX cells, as it results in the so-called "Chaotic Counting": the choice of the inactive X is still random [44], but in a fraction of cells both X chromosomes are inactivated [45].

DNA insertions in autosomes (non-sex chromosomes) including multiple copies of parts of these regions have been also analyzed. Intriguingly, in a fraction of cases, they cause the inactivation of the only X in male cells [32, 48].

Also the phenotype corresponding to the heterozygous deletion $\Delta Jpx$ is drastically different in males and females: it has no effect in the first case, whereas it is lethal in the second (at least in $\sim 85\%$ of cases), since it blocks the inactivation. The few cells which do survive the mutation, show a slight preference for the inactivation of the Wild Type X [84].

$Jpx$, like $Tsix$ and $Xite$, is a non-coding gene: it is transcribed into a RNA molecule and remains inside the nucleus. Large, non-coding RNA molecules are thought to have an important role in the inactivation process [84]. Indeed, biologists discovered that the concentration of $Jpx$ transcripts is a key factor: a reduction of the $Jpx$ transcripts, without deleting the gene itself, recapitulates the effect of the $\Delta Jpx$ deletion. Conversely, the provision of $Jpx$ transcripts in a cell carrying a heterozyogus $\Delta Jpx$ deletion, rescues lethality. Interestingly, the damages of $\Delta Jpx$ are also cancelled by a truncation of $Tsix$ gene on the X bearing the $\Delta Jpx$ deletion [84].

All these results taken together, indicate that $Xite$ and $Tsix$ act as negative regulator of $Xist$, so they tend to block the inactivation of the X chromosome; vice versa, $Jpx$ promotes the upregulation of $Xist$ gene, and so induces the inactivation of the X.

Biologists have also investigated, with a sophisticated experimental tech-

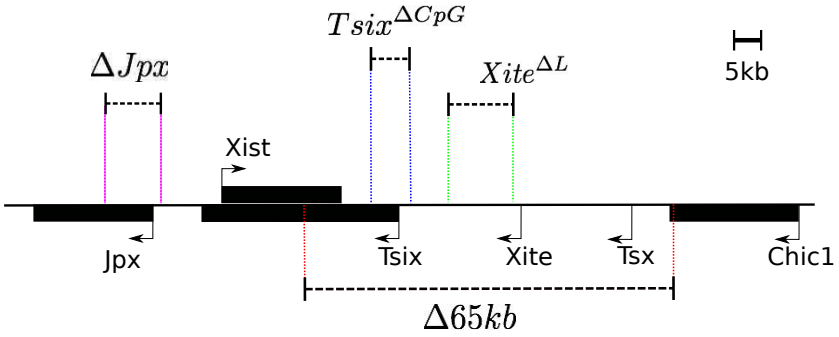


Figure 1.4: Map of some DNA deletions involving segments of the *X Inactivation Center*. See section 1.4 for a discussion on the damages they cause to X Chromosome Inactivation.

nique called "Chromosome Conformation Capture" (3C)[4] [22], the 3D spatial arrangement of the *Xic*'s during the XCI process [85]. They pointed out that the spatial architecture of *Xic* regions dynamically changes during the different stages of XCI, and is strictly correlated with the status of *Xist* gene. In particular, they were able to draw the map of genetic contacts between five specific loci within *Xic*: the promoters[5] of four genes, *Tsix*, *Xite*, *Jpx*, *Xist* and a locus close to the end of *Xist* gene which has been called "buffer" region (see fig. 1.5).

During the pre-XCI stage, DNA conformation is identical on the two X's: $Tsix$ and $Xite$ are in close proximity, folded onto the "buffer" region (see Fig. 1.5), thought to form a configuration which induces $Tsix$ transcription and, thus, $Xist$ downregulation; $Xist$ too is folded onto the "buffer" and in contact with $Jpx$, and is thus supposed to be "poised" for activation. At the onset of XCI, on the future active X, the $Jpx$-$Xist$-buffer structure opens while $Xite$ remains in contact with $Tsix$, enabling its expression and $Xist$ further downregulation (see fig. 1.5). Later on, after the Inactivation process has finished, also the $Tsix$-$Xite$-buffer structure is released and the $Xic$

[4]This technique allows to measure the relative probability that two DNA loci are close together.

[5]Region of a gene from which the transcription starts.

shows a fully opened configuration [85]. On the other X, instead, the *Tsix-Xite* interactions is lost as soon as XCI starts whereas *Xist* and *Jpx* remain in contact, resulting in *Xist* full activation [85]. The molecular factors responsible for cis-interactions across the locus are still not clearly identified, possible candidates including RNAs [46] and proteins, such as Ctcf [69, 25], which has arrays of binding sites along the locus.
The formation of higher order DNA structures is already known to have a fundamental impact on genome regulation (see section 2.2) [19, 21, 27, 39, 53, 54, 56, 83], though, how these structures are formed and controlled by the cell is still under debate [19, 18, 21, 54, 53]. In this case, it is particularly intriguing the fact that two different 3D configurations are built around the two initially identical *Xic* regions. And this happens completely at random. We will delve more into this topic in Chapter 3.

## 1.5 "Shuttling" of X chromosomes

We have seen earlier that the X's pair off at the beginning of XCI. Thereafter, when "Counting&Choice" has occurred, and the inactive X chromosome has been designated, the X's couple breaks up and the two X's reach two different nuclear targets according to their active/inactive state. Indeed, the active X chromosome is located nearby the nuclear membrane (see fig. 1.3, right panel), while the inactive X is found attached to the nucleolus[6] [94].
The attachment of the inactive X to the nucleolus is thought to be necessary for the maintenance of its silenced status [94]. Interestingly, *Xist* gene deletion on the inactive X, not only leads to a partial re-activation of the X, but also to the loss of the nucleolar targeting. Moreover, insertion on autosomes of DNA sequences including *Xist* induces autosome-nucleolus attachment, suggesting that such DNA sequences are necessary and sufficient for nucleo-

[6]Nuclear structure made of proteins and nucleic acids where the ribosomal RNA (rRNA) is transcribed and assembled. rRNA is the RNA component of the ribosomes, the machinery which builds up the proteins in the cell.

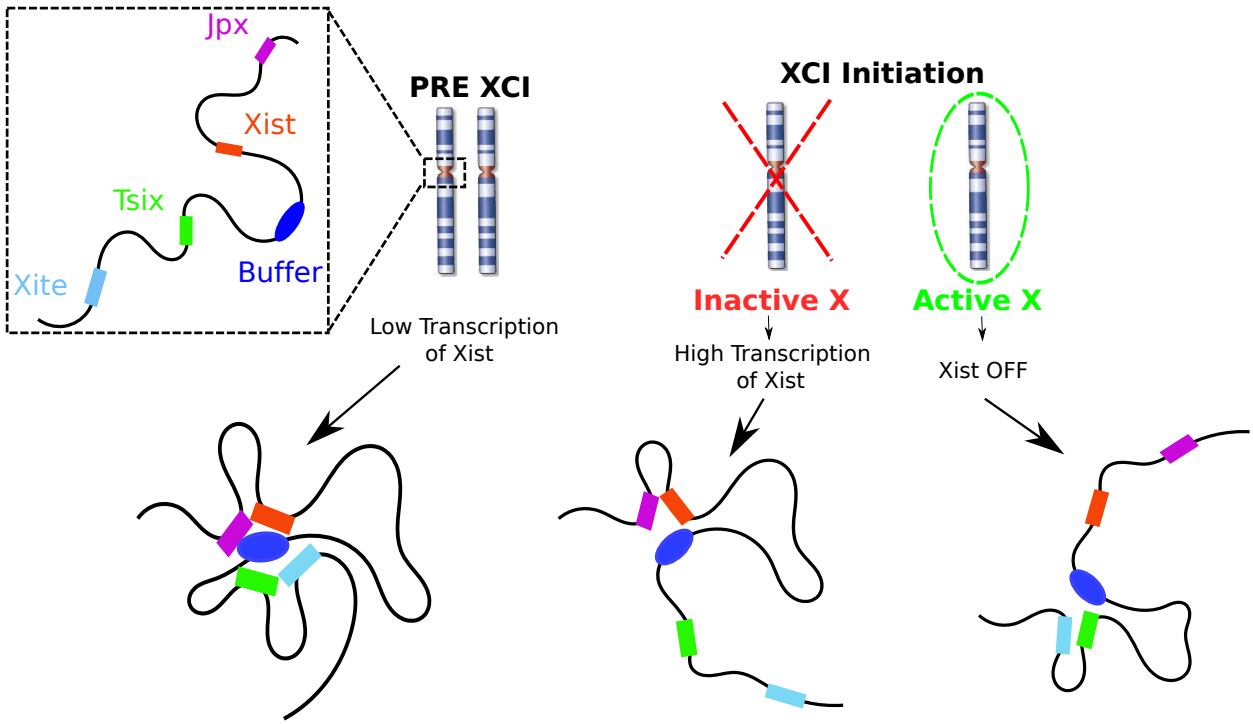


Figure 1.5: Schematic representation of the 3D configurations of DNA at *X Inactivation Center* (*Xic*) during XCI. Before XCI starts, both the *Xic*'s show the same architecture, with $Tsix$, $Xite$, $Xist$ and $Jpx$ genes attached to a region called “buffer” (coloured in blue). At XCI initiation, the active and inactive X's have different couples of genes attached to the “buffer”.

lar targeting [94].
These experimental data raise the questions on how the two X are "shuttled" to different, precise nuclear targets after XCI, and what is the exact role played by the DNA sequences which are shown to be responsible for the inactive X-nucleolus attachment.

## 1.6 X Chromosome Inactivation in multiploid cells

So far, we discussed about the XCI in diploid cells[7] having two X chromosomes, where in almost 100% of cases only one of the two X's is inactivated. Recently, XCI has been also studied in multiploid cells having more than two X's, in particular in XXXX and XXXY cells [57], where the probability distribution $\mathcal{P}$ of the number of inactive X's was measured at different times after the initiation of cell differentiation[8]. In fig. 1.6, $\mathcal{P}$ in the XXXX and XXXY cells at day 10 of differentiation is illustrated: the large majority of cells are found with 2 inactive X's in XXXX case and with 1 in XXXY. Yet, in XXXX a significant number of cells is also found with a different number of inactive X (e.g., $\sim 15\%$ of cells inactivate only one X).
Which kind of physical process can give rise to these probability distributions for the number of inactive X? Some recent works try to provide an answer to this question [72, 9, 57, 82]. We will discuss in more details this issue in Chapter 4.

[7] A diploid cell contains two copies for each chromosomes. Nearly all mammals are diploid organisms, i.e., their cells (except the sexual cells) are diploid. Cells with multiple sets of chromosomes beyond the basic set are called multiploid cells.

[8] XCI is thought to begin about two days after cells start differentiating in the embryo [93].

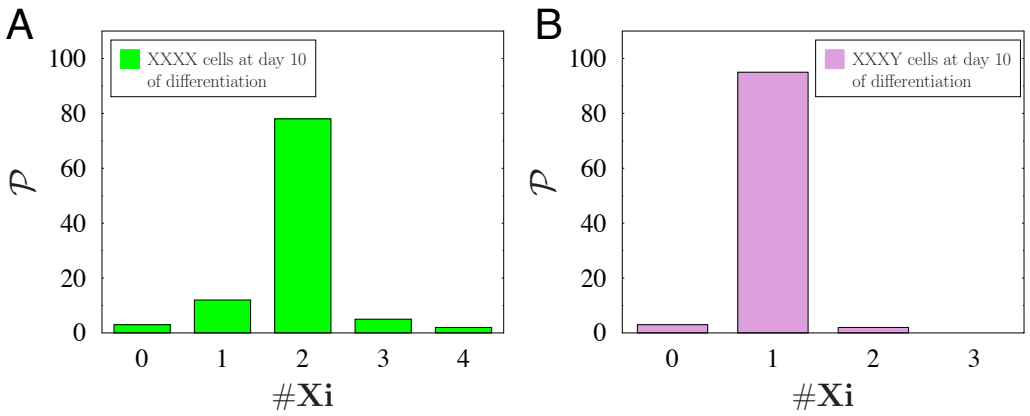


Figure 1.6: The probability distributions $\mathcal{P}$ of the number of inactive X in XXXX cells (**panel A**) and in XXXY cells (**panel B**), at day 10 of differentiation.

# Chapter 2

# "Passive Shuttling" of DNA loci

This second chapter will be devoted to the realization and the analysis of a Statistical Mechanics model for the interaction and the colocalization of a DNA locus and a target. Such a model will be proposed to explain the physical bases of the X chromosome pairing and "shuttling" that occur at X Chromosome Inactivation [73, 75, 76, 74, 9].

## 2.1 Introduction

Starting from the available experimental data about the X chromosome pairing (see section 1.3), and X chromosome "shuttling" (see section 1.5) we introduce a Statistical Mechanics model where Brownian binding molecules mediate the interactions between a DNA locus and a target, either another DNA locus (like in the X pairing) or a nuclear structure (like in the inactive X-nucleolus binding or the active X-nuclear membrane binding). In that context, we show that binding molecules can induce stable colocalization of DNA and its target via a "switch-like process", regulated by a phase transition: if the concentration/affinity of binding molecules is above a threshold value, DNA locus is reliably "shuttled" to target, despite the diffusive nature

of its motion.

Our picture can explain, thus, how well described cell strategies of upregulation of DNA binding proteins or chromatin chemical modifications can produce efficient and sharply regulated genomic architectural changes. And the scenario we depict has a close analogy with the known problem of polymer adsorption at a surface (see, e.g., [17, 20, 35, 78] and references therein). As this "shuttling process" relies on Brownian motion of DNA locus and its target, it can be seen as a form of "passive shuttling". Its robust thermodynamic roots suggest a general validity which goes well beyond the X Chromosome Inactivation (XCI).

By a mean-field theoretical approach, we describe the theoretical bases of the mechanism. Thereafter, by Monte Carlo computer simulations, we are able to quantitatively characterize the "passive shuttling" and to make predictions about the effects of possible experiments, such as DNA deletions.

## 2.2 DNA spatial architecture and genome regulation

The X chromosome pairing as well as the X "shuttling" occurring at the end of XCI (see respectively sections 1.3 and 1.5), are a prominent example of the interplay between the spatial regulation of the genome and its functioning, whose importance has been proved in a plenty of cellular processes in eukaryotic organisms[1] [19, 21, 27, 39, 53, 54, 83]. Indeed, many experiments have confirmed that DNA loci, for a correct activity, must occupy specific, but dynamically changing, positions with respect to other DNA sequences or nuclear elements. The most diverse examples of interactions exist, but the mechanisms whereby distant loci recognize each other and come together in complex space-time patterns are still largely unknown, as in the X shuttling

[1] An eukaryote is an organism whose cells are endowed with complex structures included in a nucleus which is enclosed in a membrane. E.g., all the animals and plants are eukaryotes.

case. Examples are found of loci that undergo directed motion via active, e.g., actin/myosin[2] dependent, processes [13, 26, 21, 39, 37, 80, 50]. However, most examples of DNA "cross-talks" appear to be independent of active motors. So, passive diffusion has been proposed as a major, energetically inexpensive, mechanisms [54, 19, 21]: Brownian mobility induces stochastic collisions of loci which in turn establish functional associations, e.g., via bridging molecules. Such a scenario, though, rises fundamental questions [18, 53, 64]: how are these random encounters coordinated in space and time? Are they likely? And reliable for functional purposes? How are they regulated?

Complex regulatory inter-chromosomal contacts occur, for instance, in the $\beta$-globin, $T_H2$, Hox clusters [66, 42]. Many other cases are known. The loop architecture of the Major Histocompatibility Complex class I (MHC-I) locus on human chromosome 6 [38] is mediated, for instance, by a set of specific molecules. Here chromatin loops are organized by SATB1 and PML proteins, and PML associated nuclear bodies, which tether clustered DNA binding sites to the nuclear matrix[3]. Number and position of these anchoring regions depend on the relative abundance of SATB1 and PML protein [38]: for example, while Jurkat T cells show 5 chromatin loops within such a region, CHO cells, having a lower expression of SATB1, have 6 loops which also differ in positions [38, 28]. However, if SATB1 concentration in CHO cells is matched with that of Jurkat T, a new loop organization miming that of Jurkat T cells is found [38].

Looping of specific remote loci is, for example, fundamental for the regulation of *Kit* gene in erythropoiesis (the production of red blood cells) [34]. In immature erythroid cells, where *Kit* is active, a distal $5'$ enhancer is shuttled to the *Kit* gene promoter and bound by GATA-2 proteins. Upon cell maturation, *Kit* is repressed and the above conformation changed: GATA-2

[2]Actin and myosin proteins work together to generate a molecular motor which acts, e.g., in cell division and in the movements of some organelles in the cytoplasm.

[3]Network of fibres around which chromosomes and other nuclear components are organized. Its composition and structure are still debated [2]

is displaced, while GATA-1 proteins and cofactors bring a downstream region to the promoter [34]. In this case, the relative expression level of GATA proteins acts on the chromatin conformation and controls the switch of *Kit* [34]. The list of DNA regions associated to other DNA sequences and nuclear elements is very long. Examples are the lamina-associated domains (LADs) in the human genome [31], i.e., a set of more than 1300 regions (with sizes in the $0.1 \div 10Mb$ range) found to be in close interaction with the nuclear lamina[4] (NL) and characterized by features such as a preferential occurrence of Oct-1 protein binding sites and Ctcf binding boundary elements. The reported localization of Oct-1 at the NL [33] hints to a mechanism responsible for LAD-NL binding [31]. Interestingly, clusters of binding sites are typically involved in most of the above examples[19, 53, 27, 56, 64].

It is interesting to note that all these examples, together with X pairing and X "shuttling" phenomena at XCI, share similar molecular features [19, 53, 27, 56, 64], like the need for some molecules binding to specific DNA sequences. Yet, the following questions concerning the underlying organizational principles of such complex systems remain open: how Brownian random processes can be finely regulated? How such a variety of molecular elements are orchestrated? How do they recognize each other in a distance and are brought in apposition?

## 2.3 The Model

In order to try to answer these questions, we investigate a schematic physics model of the molecular elements involved in these "shuttling events". In particular, we study two schematic models representing the situation where a DNA locus is shuttled to another DNA sequence or towards a nuclear target (e.g., nucleolus, nuclear membrane, matrix).

[4]Network of filaments which forms a thin sheetlike meshwork just beneath the inner nuclear membrane [2]. It provides mechanical support and regulates important cellular events such as DNA replication and cell division.

The DNA sequence is represented, via a standard polymer physics model, as a floating random walk polymer of $n$ beads [23]. The polymer interacts with a concentration, $c$, of Brownian molecular factors (MFs) and can be bound at a number, $n_0$, of clustered binding sites (BSs) with chemical affinity $E$ (fig. 2.1 panel A). For definiteness, here we refer to the well studied *Tsix/Xite* locus of X colocalization and choose number and chemical affinity of binding sites accordingly (see below). In real examples, the number and location of binding sites depend on the specific locus considered. However, as known in polymer physics, our thermodynamic picture is robust to parameter changes (see [23] and below).
In our model, the "nuclear target" is schematically described as an impenetrable surface having a linearly arranged set of binding sites for the DNA binding molecules (see fig,2.1 panel B). For sake of illustration, we assume that their number is also $n_0$ and their affinity $E$.

We use a simple lattice version of the random walk polymer model. This is well established in polymer physics and has the advantage to be simple enough to permit comparatively faster simulations with respect to off-lattice models. In this way we can add further degrees of freedom in our system, which represent the binding molecules, without making computation unfeasible. In fact, molecules are dealt with as a Statistical Mechanics "lattice gas" interacting with the polymer chain [81]. We consider a lattice of linear size $L_x = 2L$, $L_y = L$ and $L_z = L$ (in units of $d_0$, the characteristic size of a bead on the polymer, see below), with periodic boundary conditions to reduce boundary effects [10]. For sake of simplicity, the DNA sequence is treated as a directed polymer [23], i.e., its tips are bound to move on the top and bottom surfaces of the system volume (see Fig. 2.1). It consists of $n = L$ beads which randomly move under a "non-breaking" constraint: two proximal beads can seat only in next or nearest next neighboring lattice sites. A bond between a MF and a BS can be formed when they are on next neighboring sites; MFs can have multiple bonds (such as, e.g., Ctcf proteins). The use of directed polymers to represent DNA segments allows

faster simulations without affecting the general properties of the colocalization mechanism we describe, because it is produced by a general free energy minimization mechanism, which does not depend on such details (see below; in case of a non-directed polymer model, DNA would bind its target as well, but without a perfect alignment as in our case [64]). A strategy to attain a straight alignment anyway would be to consider a gradient of BSs along the polymer/target.

We explore these models by a Statistical Mechanics mean-field treatment and by Monte Carlo (MC) computer simulations. We try to use the available biological data to set the range of model parameters. Our models include only minimal ingredients and are very schematic, but they permit to derive a precise, quantitative picture of "passive shuttling". On the other hand, our scenario relies on a robust thermodynamic mechanism and its general aspects are, thus, not affected by the simplicity of the models.

**DNA binding sites number and chemical affinity -** Details on binding energies and DNA locations of binding sites are known in some examples (see [51, 59, 29, 43, 6, 52] and Ref.s therein), but in most cases only qualitative information is currently available.

For instance, "in vitro" measures exist [71, 70] of dissociation constants of Ctcf proteins from DNA binding sites which give binding energies around $E \sim 20kT$, $k$ being the Boltzmann constant and $T$ the room temperature (see, e.g., [95] on how to derive the binding energy from the dissociation constant). The precise value of "in vivo" binding energies depends on the specific DNA site considered and can be very hard to record, yet these "in vitro" measurements provide the typical energy range. It is experimentally well documented that DNA binding proteins, as those mentioned in the section 2.1, have a number of target loci with chemical affinities in the weak biochemical energy range, $E \sim 0 \div 20kT$ [51, 59, 29, 43, 6, 52]. This is the energy scale we consider here. Here, the BS number $n_0$ on the DNA as well as on its target is chosen to be $n_0 = 24$ (i.e., of the order of magnitude of the known case of Ctcf sites in the *Tsix/Xite* region on the X chromosome [25]), but it is varied to describe the

effects of BSs deletions.

**Molecule concentration -** The order of magnitude of the concentration of molecular factors, $c$, can be roughly estimated and compared to the concentrations of proteins in real nuclei. In our model the number of molecules per unit volume is $c/d_0^3$, where $d_0$ is the linear lattice spacing constant, which implies that the molar concentration is $\rho = c/(d_0^3 \mathcal{N}_A)$ ($\mathcal{N}_A$ is the Avogadro number). Under the assumption that a polymer bead represents a DNA segment of $\sim 20bp$ (i.e., of the order of magnitude of a Ctcf binding site in *Tsix/Xite* region [92, 25]), we obtain the order of magnitude of the polymer bead size $d_0 \sim 10nm$. By using such a value of $d_0$, typical concentrations of regulatory proteins such as $\rho \sim 10^{-3} \div 10^{-1} \mu mole/litre$ (i.e., $\sim 10^3 \div 10^5$ molecules per nucleus) would correspond to volume concentrations in our model $c \sim 10^{-4} \div 10^{-2}\%$. Such an estimate is quite rough, but may guide the connection of our study to real biological situations.

**Monte Carlo simulations -** In our Monte Carlo (MC), we run up to $10^9$ MC steps per simulation and our averages are over up to 1024 runs. In each MC step, the algorithm tries to move, on average, all the particles of the system (molecules and polymer beads, in random order), according to a transition probability proportional to $e^{-\Delta\mathcal{H}/kT}$ [10], where $\Delta\mathcal{H}$ is the energy barrier of the move ($T$ is the temperature and $k$ the Boltzmann constant). So, the binding/dissociation rate is given by the Arrhenius factor $r_0 e^{-\Delta\mathcal{H}/kT}$, where $r_0$ is the bare reaction rate. The MC time unit (i.e., a single lattice sweep) corresponds thus to a time $\tau_0 = r_0^{-1}$ (see [10]). In turn, $\tau_0$ is related to the polymer diffusion constant $D$ and to the lattice spacing constant $d_0$: $D = \left(\langle \Delta s^2 \rangle \frac{d_0^2}{4\tau_0}\right)$, where $\langle \Delta s^2 \rangle$ is the mean square displacement (expressed in units of $d_0$) of the polymer center of mass per unit MC time . While we measure $\langle \Delta s^2 \rangle$, the value of $d_0$ can be estimated to be of the order of magnitude of a typical protein binding site, say $\sim 10nm$ (see above). And we impose that the diffusion coefficient $D$ of a free polymer (i.e., with $E = 0$) in our lattice is of the order of the measured diffusion constant of human DNA loci ($D = 1\mu m^2/h$ [14]). As a result, a MC lattice sweep is found to

correspond to $\tau_0 \sim 30\mu s$ (falling well within the range of known biological kinetic constants [88]).

The above Monte Carlo moves produce an "artificial" dynamics and in general serious caution must be taken to interpret it as the real kinetics. However, in the current prevailing interpretation [10], in a system dominated by Brownian motions, a Monte Carlo Metropolis dynamics is supposed to describe well the general long time evolution of the system. Under that umbrella, we assume here that Monte Carlo could provide some insight in the system kinetics.

We consider a lattice with $L = 32$, i.e., with dimensions $L_x = 2L = 64$ and $L_y = L_z = L = 32$ in units of $d_0$. DNA segments have $n = 32$ beads. We also performed simulations with different values of $L$ and $n$ (up to $L = 128$ and $n = 128$) and we checked that our general results remain essentially unchanged. The conceptual support for using comparatively small Self-Avoiding Walk (SAW) polymer chain sizes to extrapolate the behavior of longer chains is grounded in Statistical Mechanics and rely on the system scaling properties (see [10]). For instance, the transition energy $E^*$ has a comparatively simple behavior with $n$ [64] and rapidly converges at large $n$ to a finite value comparable to $E^*(n = 32)$. Those remarks support the idea that our results are not an artifact of the specific length of the polymer.

## 2.4 Mean-Field Theory

To describe the concept behind passive shuttling and colocalization, we briefly discuss the Statistical Mechanics of the system at the level of a mean-field, coarse grained, approximation [81]. We refer to the polymer adsorption literature for more advanced theoretical approaches (see, e.g., [17, 35, 20, 78] and ref.s therein). For sake of definiteness, we consider the case with two DNA polymers (such as in the case of X Chromosome Pairing). We partition the nucleus in two halves, and name $x$ the probability to find polymer 1 in the "right half" and $y$ the probability to find polymer 2 in the "left-

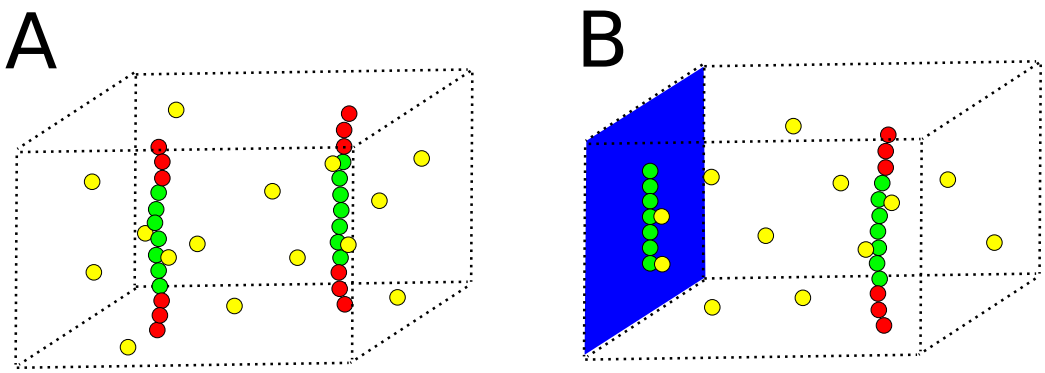


Figure 2.1: Schematic representation of the lattice models we use in Monte Carlo simulations. **Panel A** Two DNA segments are modelled as a couple of random walk polymers, endowed with a number of binding sites (green beads) for diffusing molecules (yellow beads). **Panel B** In order to study the DNA "shuttling" to a nuclear target, one copy of the polymer is replaced by a fixed, impenetrable surface (colored in blue) having the same number of binding sites (green beads).

half" (see fig. 2.2). In a Ginzburg-Landau approach [41, 81] the system Free Energy density can be written as a function of $x$ and $y$, $F \simeq F(x,y) = H(x,y) - TS(x,y)$ ($T$ is the temperature and, below, $k$ the Boltzmann constant). The interaction energy density, $H$, can be expanded in powers of $x$ and $y$ to consider the first non trivial terms:

$$H = -E_b \left[x(1-y) + y(1-x)\right] \tag{2.1}$$

The above quadratic form arises since a molecular bridge between the polymers can be formed only if they are in the same part of the nucleus. $E_b$ is the average binding energy density.
In turn, the entropy, $S(x,y)$, in such a mean-field approach can be approximated as the sum of the entropies of the two non-interacting polymers, $S(x,y) = S(x) + S(y)$, where $S(x)$ has the standard expression [81]:

$$S(x) = -k \left[x \ln(x) + (1-x) \ln(1-x)\right] \tag{2.2}$$

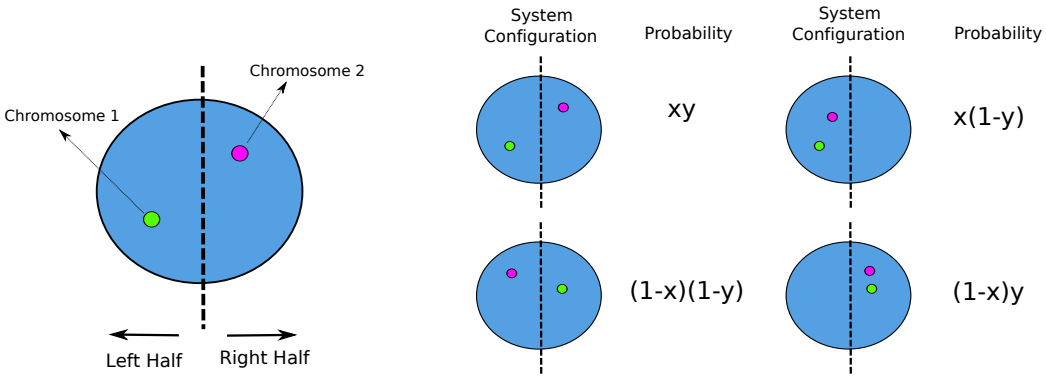


Figure 2.2: A coarse-grained version of the system is analysed via a mean-field theoretical approach. The nucleus is divided into two halves: $x$ is the probability to find the chromosome 1 (green circle) in the left half, $y$ the probability to find chromosome 2 (purple circle) in the right half. The system configurations with the relative probabilities are shown.

Thus the free energy density is:

$$F(x,y) = -E_b\left[x+y-2xy\right] + kT\left[x\ln x + (1-x)\ln(1-x) + y\ln y + (1-y)\ln(1-y)\right] \tag{2.3}$$

Note some symmetries of the function $F$:

1. $F(x,y) = F(y,x)$, namely, nothing changes if chromosome 1 and 2 are exchanged;

2. $F(x,y) = F(1-x,1-y)$, i.e., the system is symmetric under "left-right" inversion.

In the thermodynamic limit, the equilibrium states of the system correspond to the minima of $F$. To find them, let's calculate first the stationary points of $F$:

$$\begin{cases} \frac{\partial F}{\partial x} = -E_b(1-2y) + kT\ln\left[\frac{x}{(1-x)}\right] = 0 \\ \\ \frac{\partial F}{\partial y} = -E_b(1-2x) + kT\ln\left[\frac{y}{(1-y)}\right] = 0 \end{cases} \tag{2.4}$$

If we obtain $y$ from the first equation, the second equation gives:

$$x = \frac{1}{2} - \frac{kT}{2E_b} \ln \left[ \frac{1 - (kT/E_b) \ln \left( \frac{x}{1-x} \right)}{1 + (kT/E_b) \ln \left( \frac{x}{1-x} \right)} \right] \tag{2.5}$$

The graphical study of the equation (2.5) will give us the solutions of the system (2.4). The plots in fig. 2.3 show that for low values of $E_b$ (for instance, $E_b = 1.5kT$, panel A), only the symmetric solution, $(x = 1/2, y = 1/2)$, exists. Yet, as $E_b$ is increased ($E_b = 2.5kT$, panel B), two other solutions appear, $x_1^* \equiv x^* < 1/2$ and $x_2^* = 1 - x^* > 1/2$ (note that $x_2^* = 1 - x_1^*$, for the symmetry number 2, see above). While $x^*$ can be evaluated numerically, we can easily calculate the corresponding values of $y$ by using the symmetry 1 of $F$ (see above), and yield the solutions $(x, y)$ of the system (2.4): $\{(x^*, 1 - x^*), (1 - x^*, x^*)\}$.

What do these solutions physically mean? This can be easily understood by introducing the excess colocalization probability of the two polymers:

$$P(x, y) = x(1 - y) + y(1 - x) - [xy + (1 - x)(1 - y)] \tag{2.6}$$

namely, the probability to have the two polymers in the same half of the system, $x(1 - y) + y(1 - x)$, minus the probability to have them in two different halves, $xy + (1 - x)(1 - y)$ (see fig. 2.2). It can be verified that $P(1/2, 1/2) = 0$, and so that the symmetric solution, as it was easy to guess, corresponds to the case where it is equally likely to have the polymers in the same half or in opposite halves. On the contrary, with the other two solutions, $P(x^*, 1 - x^*) = P(1 - x^*, x^*) > 0$, indicating that the polymers tend to stay in the same half.

As it is seen from the plots in fig. 2.3, the minimum value of $E_b$ at which the other two solutions appear can be calculated by imposing that the derivative of the right-hand side of equation (2.5), at the point $x = 1/2$, is smaller than 1:

$$\frac{d}{dx} \left\{ \frac{1}{2} - \frac{kT}{2E_b} \ln \left[ \frac{1 - (kT/E_b) \ln \left( \frac{x}{1-x} \right)}{1 + (kT/E_b) \ln \left( \frac{x}{1-x} \right)} \right] \right\}_{x=1/2} < 1 \tag{2.7}$$

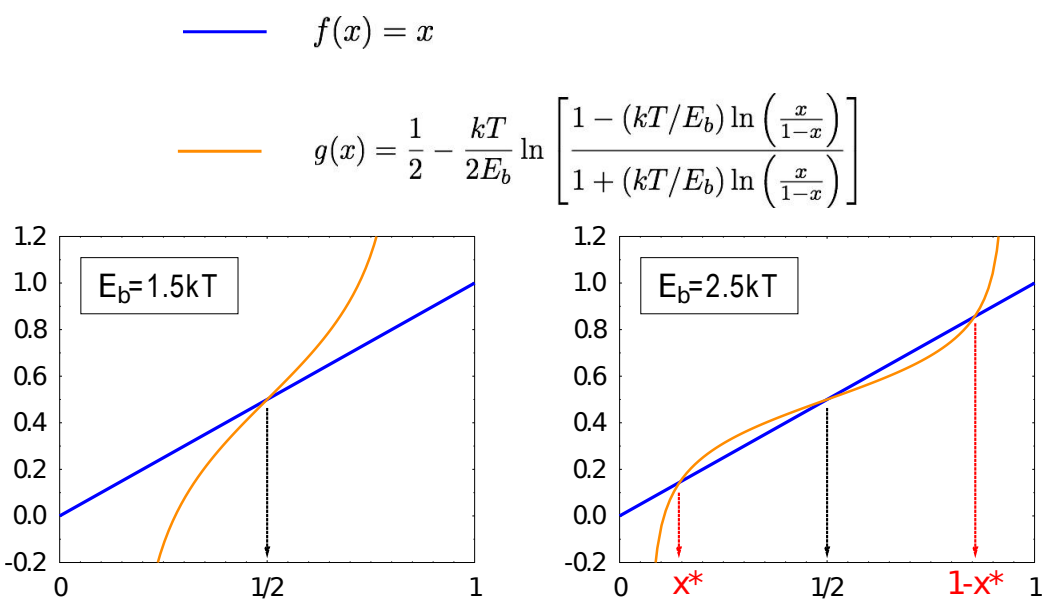


Figure 2.3: The equation (2.5) is solved graphically. The solutions are represented by the intersections of the curves $f(x)$ (blue curve) and $g(x)$ (orange curve), that are respectively the left and the right side of the equation. For $E_b = 1.5kT$ (left plot), one solution is found at $x = 1/2$, whereas, when $E_b = 2.5kT$ (right plot), other two solutions appear (red arrows).

By solving this equation, it is obtained that the solutions $x_1^*, x_2^* \neq 1/2$ are found as soon as:

$$E_b > E_b^* = 2kT \tag{2.8}$$

It can be verified that the solution $x = 1/2$ corresponds to a minimum of the free energy density $F$ if $E_b < 2kT$; conversely, if $E_b$ is larger that $2kT$, $x = 1/2$ turns into a saddle point, whereas $x_1^*, x_2^*$ are two minima.
Thus, at the thermodynamic equilibrium, if $E_b < E_b^* = 2kT$, $P = 0$ and polymers are independently located in the nucleus; above $E_b$, $P > 0$ and they are more likely to be found together in the same area (see fig. 2.4 panel A). Indeed, at $E_b^*$, a second order phase transition occurs with a consequent spontaneous symmetry breaking [81], and the excess colocalization probability $P$, is the order parameter of the system.
The critical energy value, $E_b^*$, corresponds to the point where the entropy loss owing to colocalization is compensated by the consequent energy gain.
By expanding the equilibrium value of $P$ near the critical point $E_b$, it is yielded:

$$P(E_b) \sim (E_b - E_b^*) \tag{2.9}$$

and so a critical exponent equal to 1 is found.
So far, we carried out our analysis in terms of $E_b$, the average binding energy density. Let's now go back to the system parameters we introduced in the previous section, i.e., the concentration of molecules $c$, their DNA affinity, $E$, and the number of binding sites, $n_0$. For low values of $c$ and $E$, we can approximate $E_b$ as the product of the density of available binding sites bound by a molecule, $cn_0$, multiplied by the total chemical affinity of a bridge, $2E$: $E_b(c, E, n_0) \propto 2Ecn_0$. Under this approximation, we can provide the following expression for the transition surface (see fig. 2.4 panel B):

$$cn_0E/kT = 1 \tag{2.10}$$

The advantage of the above mean-field description is to illustrate the basic ideas of the scenario we propose. However, it is very schematic and in the following sections we discuss a detailed Monte Carlo simulation of the model.

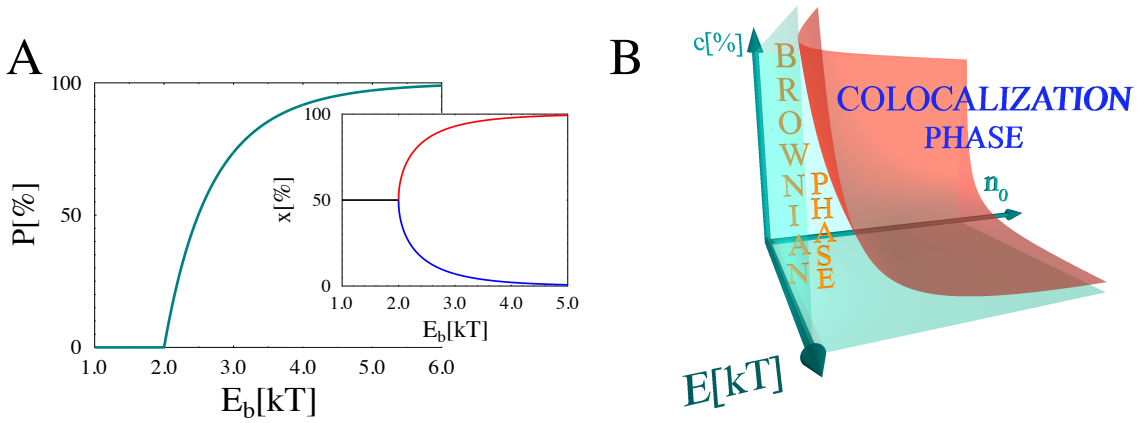


Figure 2.4: These figures illustrate the mean-field theory description of DNA-target colocalization. **Panel A** The polymer excess colocalization probability, $P$, is plotted as a function of the average binding energy density $E_b$ (The **Inset** shows the probability $x$, to find polymer 1 in the right half of the nucleus, see also fig. 2.2). The plots show how at $E_b/kT = 2$ a transition occurs between a phase where polymers are independently located in space ($P = 0\%$, $x = 50\%$) and a phase where they colocalize ($P > 0\%$, $x \neq 50\%$). In **panel B** the transition surface $cn_0E_b/kT = 1$ is depicted in the space of molecule concentration, $c$, binding energy, $E$, and number of binding sites $n_0$.

## 2.5 Monte Carlo simulations

### 2.5.1 Chromosome Pairing

Let's consider first the case where we have a couple of polymers in the lattice, representing, e.g., the two *Xic* segments in a female mammal cells (see fig. 2.1 panel A).

#### 2.5.1.1 System Thermodynamics

Since the diffusing molecules can bind both the polymers, they can induce an effective attraction force between them via the formation of molecular bridges. Though, the system reacts to a change in one of these parameters only if a threshold value in the Molecular Factor (MF) concentration $c$ and their polymer affinity $E$ is crossed: below this threshold, polymers are independent, above they colocalize, as a result of a thermodynamic phase transition [73, 61, 60].
We illustrate such an effect by considering the mean square distance $d^2$ between the binding sites (BSs) of the polymer sequences at the equilibrium:

$$d^2 = \frac{1}{n_0} \sum_{z=1}^{n_0} \frac{\langle r^2(z) \rangle}{r^2_{rand}} \tag{2.11}$$

where $r^2(z)$ is the square distance between two BSs at height $z$, averaged over all the $n_0$ BSs and over different Monte Carlo (MC) simulations (indicated by $\langle ... \rangle$). We use as a normalization constant the mean square distance between two randomly diffusing polymers, $r^2_{rand}$.
We also measure the fraction of colocalized polymers $p$, namely, the fraction of polymer couples whose equilibrium mean distance is less than 10% of the linear size of the including volume, $L$.
In presence of a given concentration, $c$, of MFs, having an affinity $E$ for the polymers, a bridge between the two polymers could be formed by chance when a MF binds both simultaneously. As a single bridging event is statistically unlikely and short lived, the degree of pairing between the polymers is

expected to be stronger the higher $c$ and $E$. However, a threshold behaviour exists. This is shown in fig. 2.5, where the equilibrium values of $d^2$ and $p$ are plotted as function of $E$ (here we fixed $c = 0.2\%$ and $n_0 = 24$): while for $E < E^* = 1.75kT$ the random values of $d^2$ and $p$ were measured ($d^2 = 100\%$, $p = 0$), practically irrespective of $E$, as soon as $E \geq E^*$, $d^2$ suddenly falls down to 0%, and, correspondingly, $p$ increases to 100%. In the crossover region around $E^*$ (defined by the criterion $p(E^* = 50\%)$ ), intermediate vales for $p$ and $d^2$ are found, since polymer couples are continuously formed and disrupted.

We found that a Fermi function well fits the simulations data for $d^2(E)$ and $p(E)$ (continuous grey lines in fig. 2.5):

$$d^2(E) = \frac{1}{1 + e^{\left(\frac{E-E^*}{\Delta E}\right)}} \quad , \quad p(E) = \frac{1}{1 + e^{-\left(\frac{E-E^*}{\Delta E}\right)}} \tag{2.12}$$

with $\Delta E = 0.03kT$ representing the width of the crossover region, where $d^2$ and $p$ are between 0 and 1.

As we explain at the beginning of this section, in the thermodynamic limit (i.e., with an infinitely large system), $E^*$ would mark the transition point between two phases (a "Brownian Phase" below it, a "Colocalization Phase" above), and $\Delta E \to 0$. As expected [81], a power-law well reproduces the two order parameters around the transition point $E^*$ (green and red dashed lines in fig. 2.5): $d^2(E) \sim -E^\alpha$, $p(E) \sim E^\alpha$ with $\alpha = 0.43$

The same non-linear behaviour of $p$ and $d^2$ is found as function of the concentration $c$ at a fixed value of the energy (see insets in fig. 2.5, here $E = 1.2kT$). So, the route to colocalization could be taken either by increasing $E$ (e.g., with modifications of the DNA binding regions or of mediating molecular complexes) or by increasing $c$ (with an increased production of MFs). This is summarized in the system phase diagram illustrated in fig. 2.6, showing the equilibrium pairing state of the polymers in the concentration-energy plane $(c, E)$, along with the transition line between the two phases (yellow line in fig. 2.6). The existence of a threshold line, $c(E)$, has its roots in a thermodynamic phase transition occurring in the system, where the energy

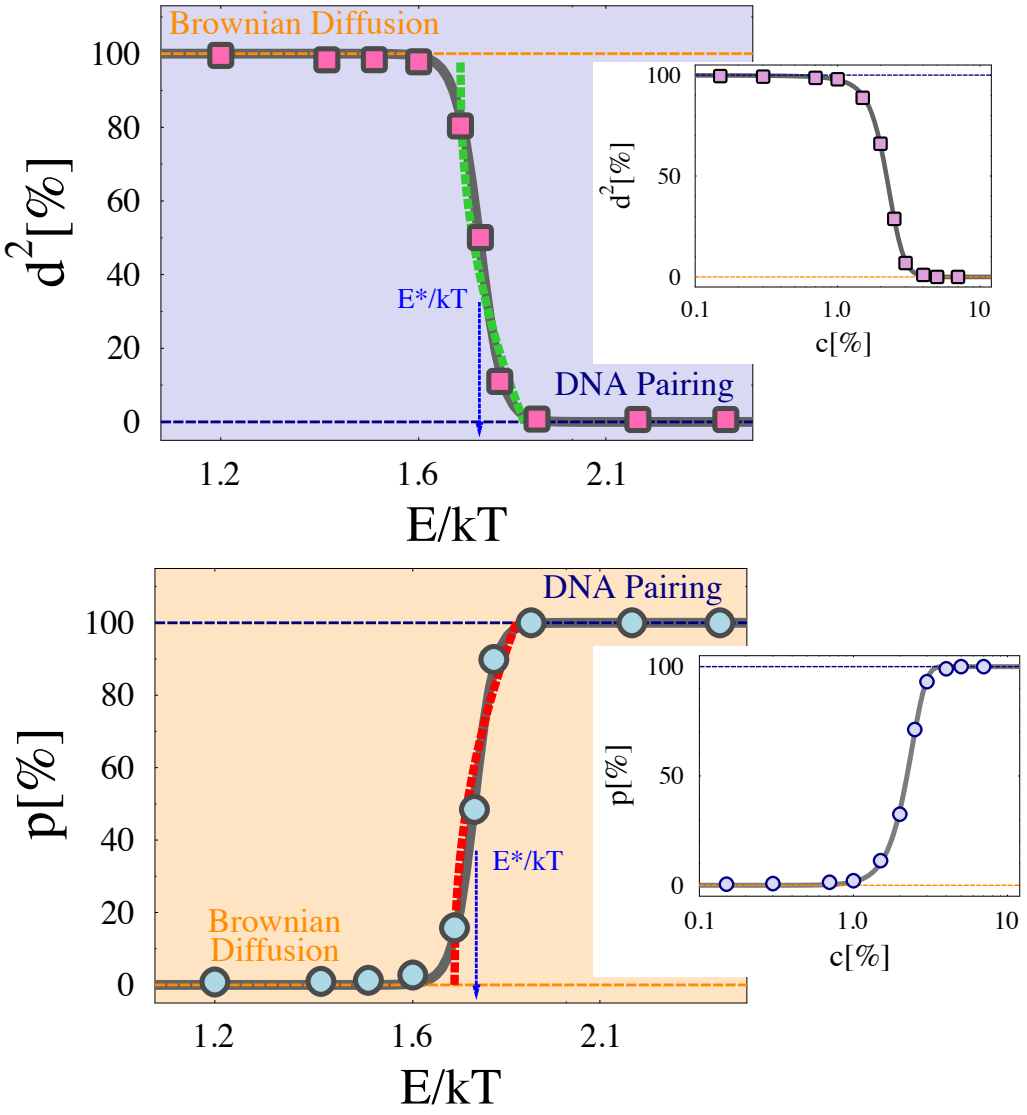


Figure 2.5: DNA loci colocalize as a result of a thermodynamic phase transition. We show this through the plots of the normalized mean square distance ($d^2$, **panel A**), and of the fraction of paired DNA loci ($p$, **panel B**) at equilibrium, as functions of $E$, the interaction energy between the molecular factors (MFs) and the DNA binding sites (BSs). We took a MF concentration $c = 0.2\%$ and a DNA BS number $n_0 = 24$. As $E$ increases, $d^2$ and $p$, from values typical of Brownian diffusion ($d^2 \sim 100\%$ and $p \sim 0\%$, orange horizontal lines), rapidly saturate to values which indicate DNA loci full colocalization ($d^2 \sim 0\%$, $p \sim 100\%$, blue horizontal lines), after a threshold value $E^*/kT = 1.75$ (defined by the criterion $p(E^*) = 50\%$). Superimposed fits for both $d^2(E)$ and $p(E)$ are Fermi functions (grey lines, see eq.s (2.12) ). Around the transition point $E^*$ (blue arrow), a power-law fits well (green and ref dashed lines). $d^2$ and $p$ show the very same behaviour as function of the molecular factor concentration $c$ (see **insets**; here $E = 1.2kT$ and $n_0 = 24$).

gain resulting from pairing compensates the corresponding entropy loss, as we point out in the mean-field theory (section 2.4). The transition line $c(E)$ at small $E$ is well fitted by an inverse power-law :

$$c \sim (E - E_{min})^{-\nu} \tag{2.13}$$

where the exponent is $\nu \sim 4$ and $E_{min} \sim 0.7kT$ is a minimal threshold energy below which no pairing transition is possible. At higher values of $E$ an exponential fit works as well.
This phase diagram gives precise constraints to the admissible values of $c$ and $E$ to attain colocalization. On the other hand, colocalization is found in a broad range of biologically relevant values for $c$ and $E$ (see section above page 19). This confirms that the present colocalization mechanism is robust and does not depend on the biochemical details of the system.

#### 2.5.1.2 Dynamics of pairing

The mechanism of "passive shuttling" we are describing leads to stable binding between the two polymers, but it must be also fast enough to serve functional purposes. We investigate its dynamics as produced by MC simulations, a very schematic description, yet considered to well describe the general long time evolution of a system dominated by Brownian processes (see above page 20 and [10]).
We measured the time behaviour of the mean square distance $d^2(t)$ (see eq.(2.11)), which is plotted in fig. 2.7 (upper panels) for two values of $E$, one in the "Pairing Phase", the other in the "Brownian Phase" (see previous section). At $t = 0$, the MFs were randomly placed within the lattice sites, while the polymers were in their straight vertical configuration, at a distance $L$ one from each other (so, the starting value of the distance is $d^2(t = 0) = L^2/r^2_{rand} \sim 2.5$). As expected from the results of the previous sections, the increase of $E$ determines very different equilibrium states for the system: while at lower energy, $d^2$ saturates at the value expected from a couple of independent random walk polymers ($d^2 \sim 100\%$, see upper left

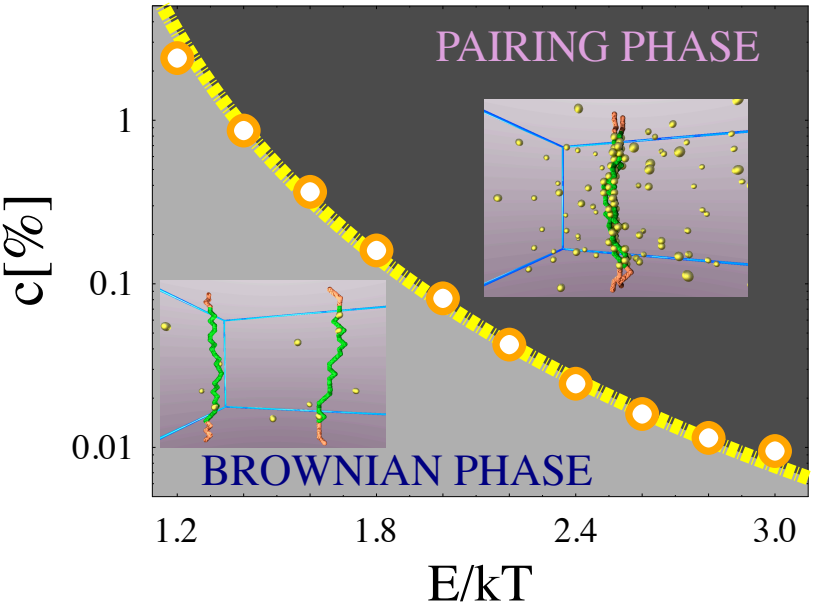


Figure 2.6: The diagram shows the thermodynamic equilibrium state of the system in the $(E, c)$ plane, for a range of typical biochemical binding energies, $E$, and molecular concentrations, $c$. Circles mark the transition points from the Brownian phase, where chromosomes diffuse independently, to the Pairing phase, where chromosomes are juxtaposed (the superimposed fit is a power law). Two typical equilibrium configurations of the system in the Brownian and in the Pairing phase, generated by Monte Carlo computer simulations, are shown.

panel, red vertical dashed line), when $E = 2.2kT$, $d^2$ eventually goes down to $\sim 0$ (upper right panel, red vertical dashed line), thus revealing that the effective attraction is now high enough to make DNA loci pair off.
$d^2(t)$ shows an initial linear behaviour, determined by the random diffusion of the polymers when they are far from each other, and a final exponential saturation to a plateau which is dependent, as we saw, on the values of $E$ and $c$. A good fit function, which includes both the above mentioned time regimes, is the following (gray lines in fig. 2.7 upper panels):

$$d^2(t) = d^2(\infty) + \left[ d^2(0) - d^2(\infty) + \frac{at}{1+bt} \right] e^{-\frac{t}{\tau}} \tag{2.14}$$

where $d^2(\infty)$ is the final equilibrium value, while $a$, $b$ and $\tau$ are fit parameters which depend on $E$ and $c$.
To get more insight into the dynamics of this pairing mechanism and into the key role of the molecular factors, we monitored the time behaviour of the average fraction $F$ of the MFs attached to a single DNA locus. The role of MFs as mediators of the DNA loci interaction clearly emerges in the middle panels in fig. 2.7, where the plot of $F(t)$, for the same values of the parameters used above is shown.
For both the energies $E$, $F(t)$ starts from the same value, due to the initial random distribution of the MFs in the lattice sites (compare right and left middle panels). When $E = 1.4kT$, $F(t)$ simply saturates in a time scale which is about 4 orders of magnitude smaller than that of $d^2(t)$ (green vertical dashed line in left panels); indeed, MFs have a size much smaller than polymers, so their dynamics is much faster. The molecules remain bound to a single DNA locus, without forming any "bridge" at all times (left lower panel). If the energy is increased to $E = 2.2kT$, and the DNA loci pair off ($d^2$ saturates at $\sim 0$), a more interesting time evolution is found. In fact, now, two time regimes in $F(t)$ are easily distinguished: during the first $\sim 0.5min$, $F(t)$ has a behaviour similar to that in the previous case, except for the higher plateau due to the higher MF-BS affinity (green vertical dashed line in right panels). This first time regime corresponds to the initial MF binding to the

DNA loci which, still out of the action range of the MF-induced effective potential, independently diffuse. Yet, just when $d^2$ begins to exponentially fall down to $\sim 0$, $F(t)$ rises to a second plateau (red vertical dashed line): as soon as DNA loci, during their random diffusion, get closer, some MFs start to stably bind both of them, and, as a result, $F(t)$ increases. This is evidenced in the lower right panel of fig. 2.7, showing that quite all the bound MFs actually form "bridges" between the DNA loci after $\sim 1min$. These MFs "bridges" keep together the two DNA loci, so that pairing is finally produced. The $F(t)$ analysis, while allows an immmediate distinction between two dynamical regimes, makes the MF mediating role evident: we saw that colocalization takes place as a result of MF bridges realization, which, however, occurs at thermodynamic equilibrium only for certain values of the parameters.

The parameter $\tau$ in the fit of $d^2$ (see eq. (2.14)) represents the characteristic time needed to reach the equilibrium distance and pairing (in the "Pairing Phase"). We studied it as function of $E$ and $c$ (fig. 2.8). Interestingly, an increasing behaviour is found both for $\tau(E)$ (fig. 2.8 main panel) and $\tau(c)$ (inset in fig. 2.8), with a "jump" at the transition (see fig. 2.8 blue arrows).

For fitting (grey and red dashed lines in fig. 2.8) we used power-law functions, as this is the expected behaviour for the relaxation time near a transition point, in the thermodynamic limit [81]. However, in our finite-sized case, we found that exponential functions can fit the data as well.

The rationale for such an increasing behaviour of $\tau$ is that when DNA segments are bound by MFs their effective diffusion constant is proportionally reduced (see also below) and so, while they will be eventually colocalized, the time to reach equilibrium is longer the higher the concentration and the DNA affinity of MFs. This could result in constraints on the biologically admissible values of $E$ and $c$ since, even if a stable DNA pairing is eventually achieved, the whole process could take too long if $(c, E)$ are too high. It is also interesting to note how, while only reasonable guess values are used for the system

parameters (i.e., molecular concentration/affinity, number of binding sites, kinetic rate constants, etc., see page 19), the values of $\tau$ predicted by Monte Carlo simulations in the "Pairing Phase" are of the order of magnitude of the characteristic time scales of cellular processes ($\sim 10 \div 10^2\, min$ see [88]).

As we already stressed, the dynamics of colocalization is interesting in itself and still experimentally largely unexplored. For this reason, we went further by measuring the mean square displacement from the initial position, $\langle \Delta s^2 \rangle$, of the polymer center of mass as function of time. In fact, this is a main quantity describing the kinetics, as the dynamics of polymers is diffusive in nature.

$\langle \Delta s^2 \rangle(t)$ of the centers of mass of the DNA segments is plotted for two values of the energy $E$ (one in the "Brownian Phase", one in the "Pairing Phase", see above) in fig. 2.9, panel A. Note that the linear behavior of $\langle \Delta s^2 \rangle(t)$, typical in diffusive regimes, is observed at short as well as at long times (grey dashed lines), for both the values of the energy. To characterize these diffusive regimes, we measured the short and long time diffusion constants $D = \frac{\langle \Delta s^2 \rangle(t)}{4t}$ (named respectively $D_0$ and $D_\infty$) for the two values of energies. At low energy (e.g., $E/kT = 1.4$), when the colocalization machinery is off, Brownian motion has approximately the same diffusion constant at short and long times. At higher energy (e.g., $E/kT = 1.9$) two distinct diffusive regimes are found: an initial one when the two polymers diffuse independently, and a longer time slower diffusion when they move bound to each other. Such a behavior is captured by a plot of the short and long times diffusion constants, $D_0$ and $D_\infty$, as function of $E$ (see fig. 2.9 panel A inset). As seen before, $D_0(E)$ decreases with $E$ and $D_\infty(E)$ follows it. The transition point, $E^*$, is marked by a drastic reduction of $D_\infty(E)$, while no major changes are found in the behavior of $D_0(E)$. Above $E^*$, $D_\infty(E)$ is non-zero since the two paired DNA segments continue to diffuse, although with a diffusion constant which is some orders of magnitude smaller than in the free case (see fig. 2.9 panel B). Such a large reduction is due to the much larger mass of the diffusing object in the "paired" state, which is formed by the couple of polymers and

by a number of attached molecules.

Note that in our model we consider only short segments of DNA. For that reason, a comparatively strong effect of $(c, E)$ on $D$ is found. In real systems, the situation is more complicated, also because chromatin[5] kinetics can be affected by other phenomena (such as chromatin entanglement, crowding, etc.) we do not consider at the level of our schematic description here. Nevertheless, a dependence of the diffusion constant on $(c, E)$ should be observed locally, at the scale of the pairing sequences.

[5]Chromatin is the combination of DNA and proteins that makes up chromosomes.

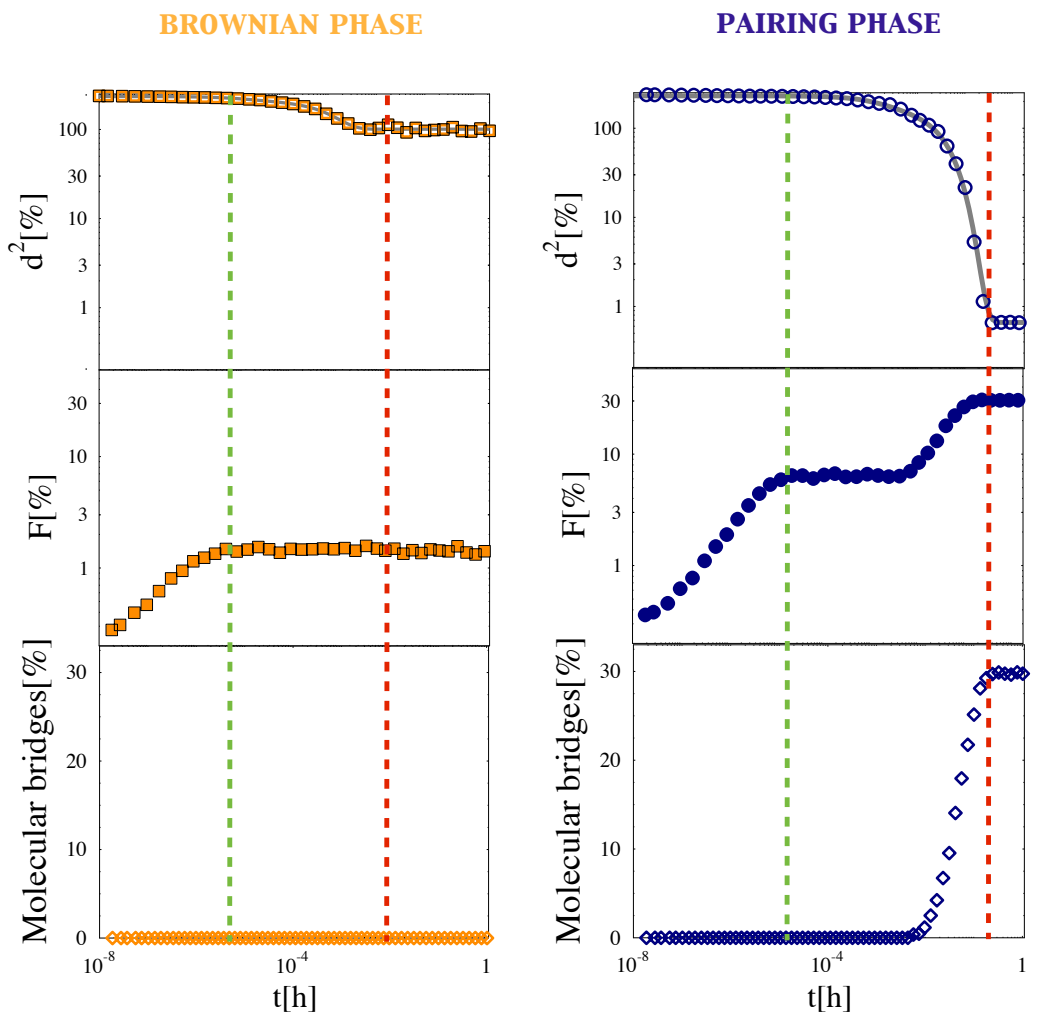


Figure 2.7: The normalized mean square distance $d^2$ (**upper panels**), the average fraction $F$ of molecular factors (MFs) bound to a single DNA locus (**middle panels**), and the percentage of molecular factors bound to both DNA loci ("molecular bridges", **lower panels**), are plotted as function of time $t$, for the system in the "Brownian Phase" ($E = 1.4kT, c = 0.2\%, n_0 = 24$; left panels) and in the "Pairing Phase" ($E = 2.2kT, c = 0.2\%, n_0 = 24$; right panels). The vertical dashed lines indicate the two time scales which characterize the system dynamics: one is the time needed to the MFs to equilibrate with the DNA loci far from each other (green line); the second one (right line) corresponds to the equilibration of $d^2$. Note that, if the system is in the "Pairing Phase" (right panel), $d^2$ saturates at $\sim 0\%$, and correspondingly $F$ rises to a second plateau as some MFs form "bridges" between the DNA loci (lower right panel).

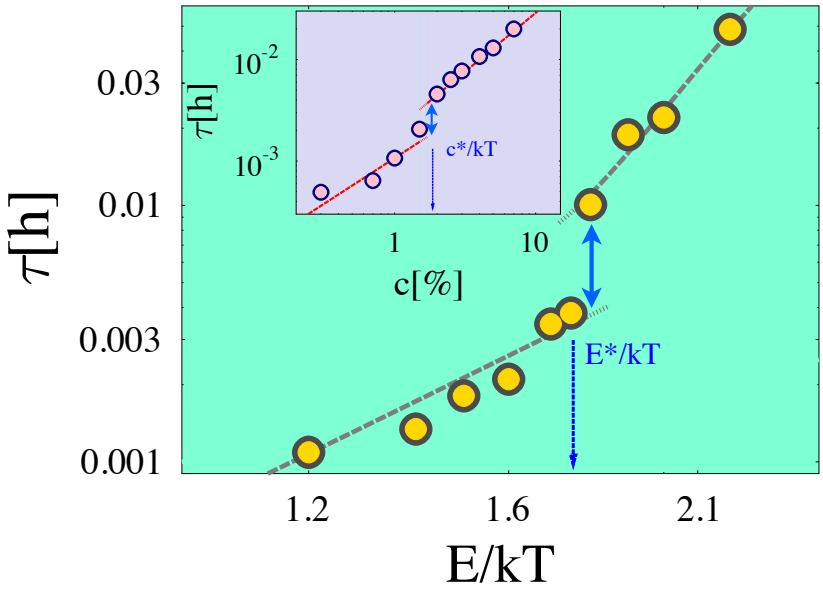


Figure 2.8: The average time scale, $\tau$, to attain the equilibrium distance is plotted as function of DNA-molecule affinity $E$ (**main panel**, $c = 0.2\%, n_0 = 24$) and of molecule concentration $c$ (**inset**, $E = 1.2kT, n_0 = 24$). $\tau(E)$ turned out to be an increasing function of $E$ as well as of $c$, as stronger and more abundant molecule-DNA bonds make the DNA loci diffusion more difficult (see also fig. 2.9). A "jump" is observed at the transition point (blue arrows). The superimposed fits are power-law functions.

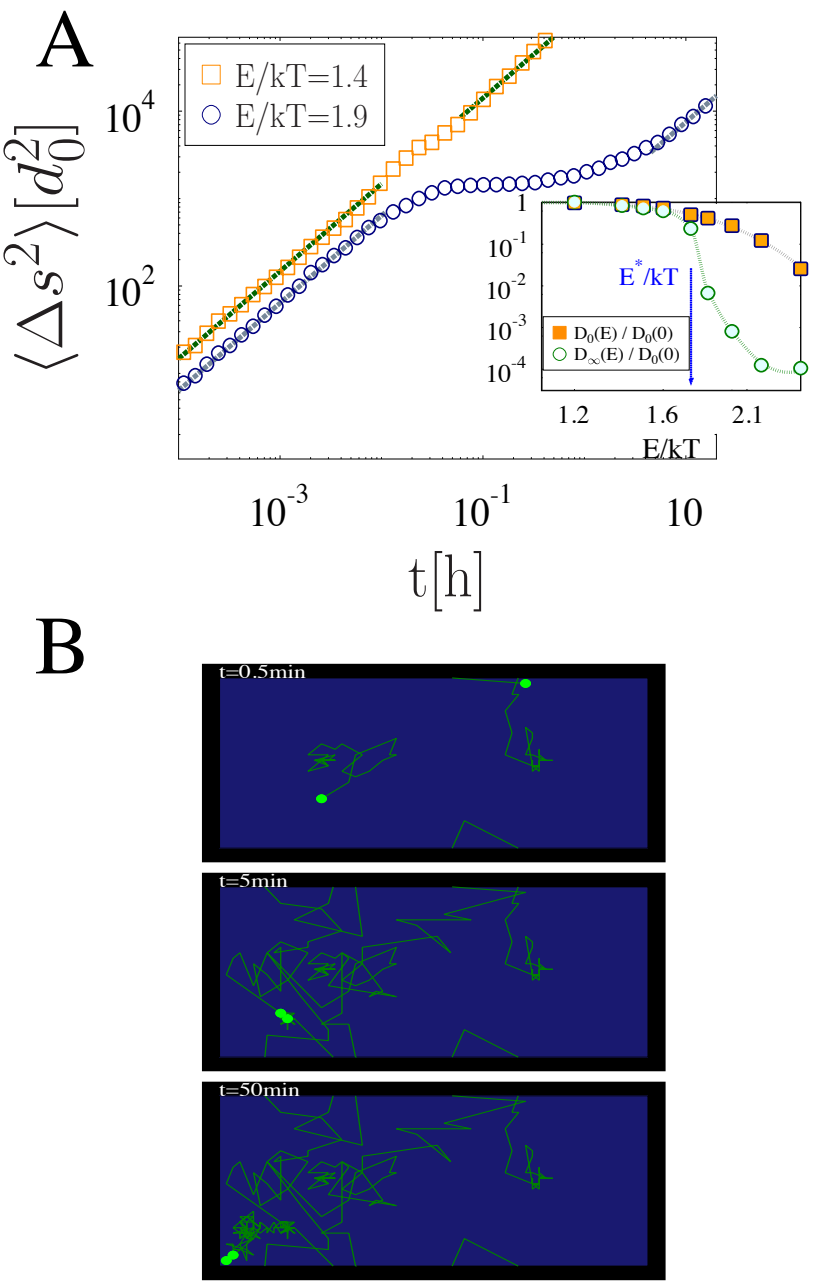


Figure 2.9: **Panel A** The mean square displacement of the center of mass of one of the DNA segments, $\langle \Delta s^2 \rangle(t)$, is plotted as a function of time $t$. At small binding energies (squares), $\langle \Delta s^2 \rangle(t)$ has a linear diffusive behavior at all times as no colocalization occurs. At higher energies (circles) two different diffusive regimes are found at short and long time scales, before and after colocalization. **Inset** The diffusion constants at short and long time scales ($D_0$ and $D_\infty$) as function of $E$, normalized by $D_0(E = 0)$. **Panel B** 2D projections of the system trajectory from a Monte Carlo simulation, showing the initial Brownian diffusion of two separated DNA loci ($t = 0.5min$) and their colocalization ($t = 5min$ and $t = 50min$).

#### 2.5.1.3 Effects of DNA deletions

We also investigated the effects on pairing of heterozygous deletions (see footnote on page 7), an issue of practical relevance to experiments. Indeed, we analyzed the system response to a variation in the number of binding sites (BSs) $n_0$ on one of the two polymers. In particular, we consider the case where a fraction $\Delta n/n_0$ of the original number of BSs is removed from one polymer (for the same MF concentration, $c = 5\%$, and affinity $E = 1.2kT$). While a reduction of the pairing fraction, $p$, is expected in presence of a deletion, we find that the equilibrium value of $p$ has a non-linear behavior in $\Delta n/n_0$. In fact, it is a sigmoid with a threshold at $\Delta n^* \sim 50\% n_0$ (see Fig. 2.10 panel A): short deletions (say, removing a fraction of BSs $\Delta n/n_0 \geq 30\%$) do not result in a relevant reduction of $p$, while pairing is completely lost as soon as $\Delta n/n_0$ gets larger than about 70%. The sigmoid behavior stems from the non trivial thermodynamic origin of the MF-mediated effective attraction between polymers.

While these results rationalize the observed length dependent effects of *Tsix/Xite* deletions (see section 1.3 on page 5 and [93]), the predicted behavior of $p(\Delta n)$ could be experimentally tested.

A Fermi fit function, as that used before (see eq. (2.12) and fig. 2.5) also works here for $p(\Delta n)$:

$$p(\Delta n) = \frac{1}{1 + e^{\left(\frac{\Delta n - \Delta n^*}{\Delta(\Delta n)}\right)}} \tag{2.15}$$

with $\Delta n^* \sim 50\% n_0$ and $\Delta(\Delta n) \sim 8\%\, n_0$.

The threshold value, $\Delta n^*$, is a decreasing function of $E$ and $c$. This is summarized in the phase diagram of fig. 2.10 panel B, showing the effects on pairing of the three parameters $(c, E, n_0)$: below the transition surface, the Brownian phase is found, above it polymers are colocalized.

Interestingly, the time to approach the equilibrium state, $\tau$, is larger the smaller the deletion, as shown in the inset of Fig. 2.10 panel A. Such a seemingly counterintuitive result stems from the fact that the smaller the BSs number, the smaller is the number of MFs attached to chromosomes and,

thus, larger their effective diffusion constant (see paragraph on page 31). The behavior of $\tau$ as function of $\Delta n$ results in a non-trivial prediction: the removal of a fraction of BSs within a chromosome should speed pairing with respect to the Wild Type case (i.e., without any DNA deletion), although, the overall fraction of paired chromosomes would be reduced. As seen in fig. 2.8, an analogous phenomenon would be observed by decreasing the concentration of MFs and/or their DNA affinity.

#### 2.5.1.4 *Xic* pairing in multiploid cells

XCI in cells with more than two X's is an important, current field of experimental investigations by biologists. Recent papers (see, e.g., [57]) point out, e.g., a non trivial probability distribution $\mathcal{P}$ of the number of inactive X in XXXX cells and XXXY cells (see page 12 and fig. 1.6), conversely to what happens in diploid XX cells, where 100% of cells have 1 inactive X. Whether or not the *Xic* pairing plays a major role in the shape of $\mathcal{P}$ is still obscure, as there are no available data about *Xic*'s pairing in multiploid cells.
By use of our model, we can make prediction about *Xic* pairing even in multiploid cells. To this aim, we considered the lattice we used before (see section on page 17) with three and four identical DNA polymers, and we sought for the thermodynamic equilibrium states of the system.
We investigated two scenarios, by making different assumptions on the DNA-molecule interaction. In a first scenario, a DNA bead can bind a single molecule at a time. Fig. 2.11 panel A shows the probability at equilibrium of each possible configuration in a system with 4X, as function of the binding energy $E$ at a fixed value of the concentration $c$. Two regimes are found: at low energies, no paired polymers are found, but for high values of $E$ the configuration with two independent polymer couples is the most likely to occur. These two regimes, which in the thermodynamic limit correspond to different phases, are separated by a narrow region where, on average, only a single though unstable polymer couple is formed in most of cases.
This result is also illustrated in fig. 2.11 panel B, where the phase diagram in

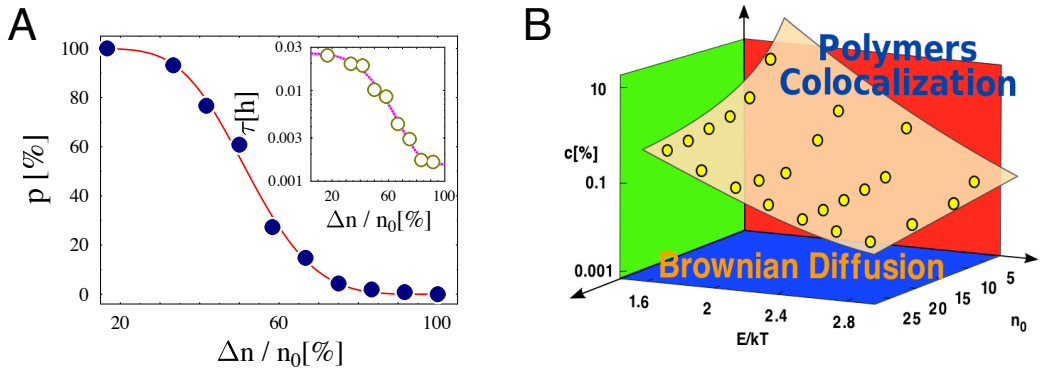


Figure 2.10: **Panel A** The figure shows the pairing probability, $p$, in heterozygous deletions, as a function of the fraction $\Delta n/n_0$ of binding sites removed. In the "Wild Type" case ($\Delta n/n_0 = 0$) the system is chosen to be in the "Pairing Phase" (here $c = 5\%, E = 1.2kT$) and the equilibrium value of the fraction of paired chromosomes is $p = 100\%$. The pairing fraction, $p$, has a non linear behavior as function of $\Delta n$, with a crossover region around $\Delta n^* = 50\% n_0$. Short deletions, say, deletions with $\Delta n \lesssim 30\%$, have tiny effects on the pairing fraction, while deletions with $\Delta n \gtrsim 70\%$ erase pairing. **Inset** The average time, $\tau$, to approach the equilibrium state is plotted as function of $\Delta n/n_0$. When the deletion gets larger, $\tau$ is shorter, since less MFs are bound to the polymers which, in turn, have an higher effective diffusion constant. **Panel B** The colocalization state of a model system with two DNA segments is summarized by this phase diagram in the $(c, E, n_0)$ space, showing the regions where pairing and Brownian, independent diffusion occur. The yellow circles mark the transition points derived from Monte Carlo simulations and the transition surface is obtained by a power law fit: $c\,(E - \tilde{E})^{\gamma}\,(n_0 - \tilde{n_0})^{\delta} = const.$, with $\gamma = 4$, $\tilde{E} = 0.8kT$, $\delta = 1.1$ and $\tilde{n_0} = 3.5$.

$(c, E)$ plane is sketched. It emphasizes how the phase transition is found in a wide range of molecule concentration and binding affinity. The red circles mark the measured transition points (defined as the inflection points of the blue and yellow curves in fig. 2.11 panel A, i.e., those corresponding to the

probability to have no paired polymers and two independent polymer couples), and the grey area represents the crossover region, which gets narrower as the thermodynamic limit is approached.
A similar picture is found in the system with 3 polymers: also in this case, the equilibrium state of the system drastically changes at certain values of the concentration/affinity of the molecules, switching between a configuration with uncoupled, independent polymer segments and a regime where all the three polymers are paired (see fig. 2.11 panel C and 2.11 panel D). An intermediate region exists with a single, unstable polymer couple (fig. 2.11 panel C purple circles and fig. 2.11 panel D grey region).

We explored a second scenario where a DNA bead is allowed to form multiple bonds with the molecules (like in the case we studied in the previous sections, for "diploid" systems)[6]. While the same results are found in the system with 3 polymers (with a slight decrease in the threshold values of $(c, E)$ needed for colocalization), an important modification of the "Colocalized State" in the 4 polymer system is observed: the phase with two stable and independent polymer couples is replaced by a phase with all the 4 polymers paired together (data not shown).

### 2.5.2 "Shuttling" to a nuclear target

The same physical mechanism which can produce the pairing of DNA sequences, can as well "shuttle" a DNA sequence to a nuclear target, like the nucleolus, the nuclear membrane (as in the case of X chromosomes at the end of XCI), etc.
To show this, we consider the variant of the model including one polymer and an impenetrable surface modelling the DNA target (see above page 17 and fig. 2.1 panel B).

[6] In the "diploid" case these two scenarios on the molecule-DNA interaction produce the same results. Indeed, if we imposed that, at the most, only one molecule can be bound to a DNA "bead", the only effect which would be noticed is an increased value of the thresholds in $(c, E, n_0)$ needed for colocalization

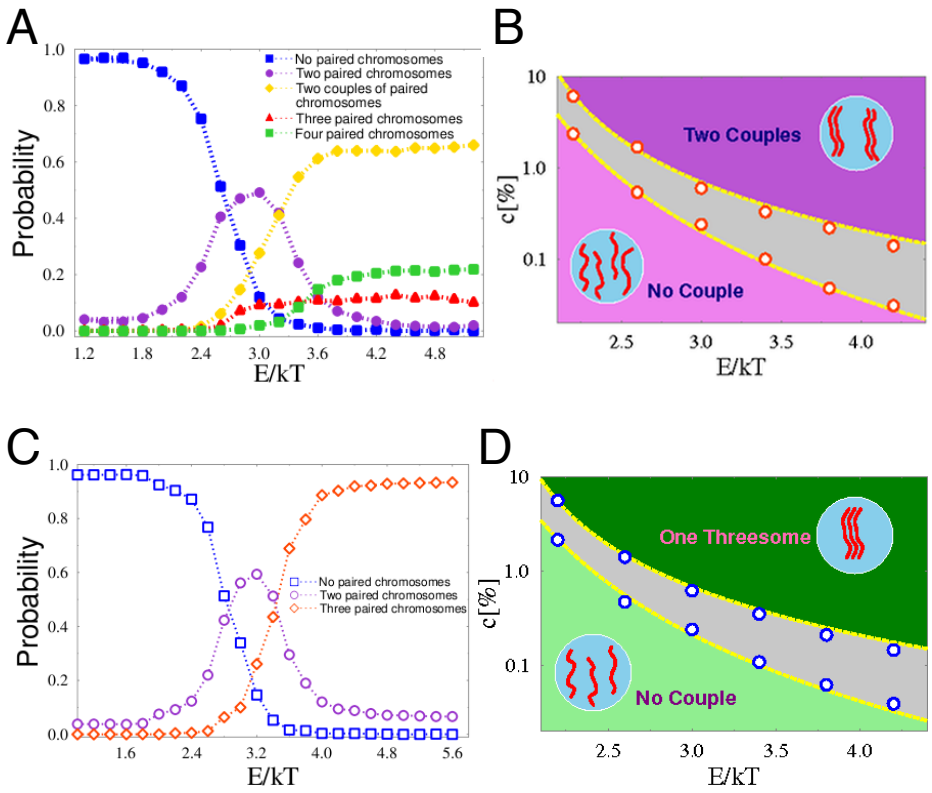


Figure 2.11: **Panel A and B** Pairing in a system with 4 polymers (such as the case of the 4 $Xic$ segments in XXXX cells). In panel A the probability of each possible configuration of the polymers is plotted as function of the molecular affinity $E$ at a fixed molecular concentration (here $c = 0.5\%$). At low energies no pairing is observed, with almost all the polymers unpaired. Conversely, at higher energies two independent polymer couples are formed in most of cases (yellow diamonds). These two regimes (respectively the "unpaired" and the "paired" regimes) are separated by a crossover region (peak of purple circles) where a single unstable couple of polymers is found with $\sim 50\%$ of probability. Panel B shows the phase diagram in the $(c, E)$ plane. The red circles mark the transition points, the dashed yellow line are power-law fits. The grey area represents the crossover region with just one unstable couple formed. **Panel C and D** refer to a system with 3 polymers. In panel C the probability of the different pairing configurations at equilibrium is plotted as function of $E$ (with $c$ =0.5%). Two thermodynamic phases can be seen: no X couples at low $E$, while all the 3 X are paired above a threshold (after a narrow crossover region). The phase diagram in $(c, E)$ plane is illustrated in panel D, showing the values of $(c, E)$ (blue circles) at which the phase transition occurs (the crossover region is in grey). Superimposed fits (yellow dashed lines) are power-laws.

#### 2.5.2.1 Thermodynamics of the system

The fig. 2.12 panel A shows a scenario which is analogous to that found for DNA pairing, with the only difference that now the target is a fixed nuclear structure. Thus, the normalized mean squared distance between the DNA and the target at equilibrium (defined in eq. (2.11)), switches from 100% to 0% with a strongly non linear behaviour as function of the DNA-molecule affinity $E$ (fig. 2.12 panel A, main panel), as a result of the thermodynamic phase transition. As before, the DNA is "shuttled" toward the target only if the binding affinity $E$, the molecular concentration $c$ and the binding sites number on the DNA, $n_0$, are above precise thresholds, as the 3D phase diagram in the inset of fig. 2.12 panel A illustrates.

#### 2.5.2.2 Role of non-specific binding sites

It is interesting to try to describe the effects on the "shuttling" mechanism of the presence of a number of unspecific binding sites on DNA/target. We discuss it in this section, in the context of DNA "shuttling" to a nuclear target, but the very same conclusions can be achieved in the case of DNA pairing.

The problem of how sequence-specific proteins can find their DNA target sites on very large eukaryotic genomes is ancient (see, e.g., [49, 87, 7]). It has been proposed that the presence of non-specific binding sites (i.e., binding sites having lower affinities) allows a mixture of one-dimensional diffusion of bridging molecules along the DNA and three-dimensional diffusion in the surrounding medium, which could result in a more efficient search of the DNA target sites than a purely one- or three-dimensional diffusion [8, 90, 89, 29, 79]. On the other hand, binding of molecules to these sites is expected to impair "shuttling" by the reduction of the effective concentration of diffusing molecular mediators.

We tested the effect of the presence of non-specific sites in our schematic

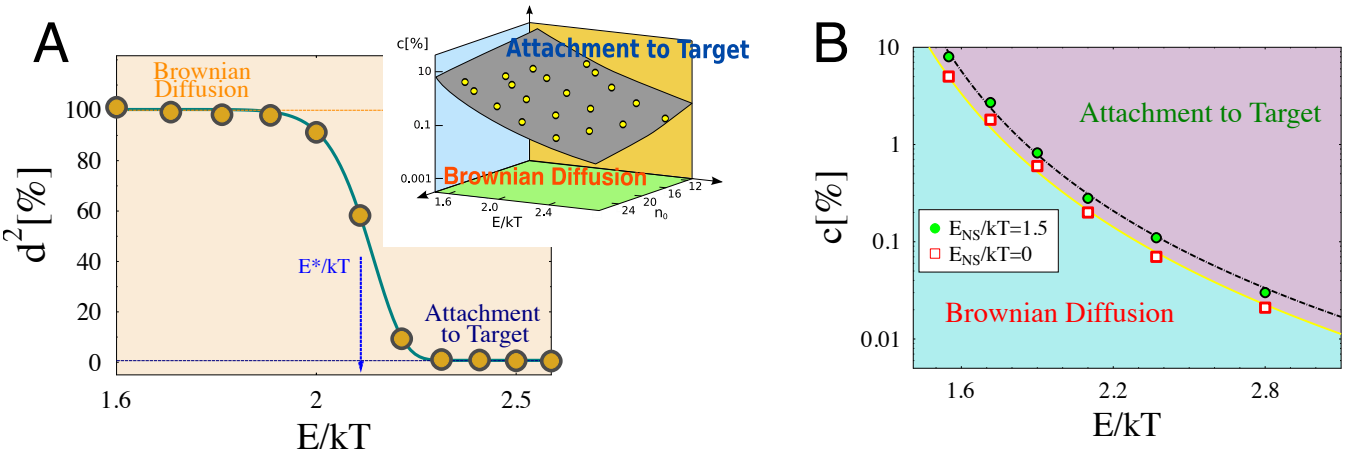


Figure 2.12: **Panel A** The equilibrium normalized square distance $d^2$ between the DNA polymer and its nuclear target is shown as a function of $E/kT$, the binding molecule affinity. At small $E$, $d^2$ has a value corresponding to random diffusion ($d^2 = 100\%$); above a threshold $E \sim E^* = 2.1kT$ (blue vertical arrow) a phase transition occurs and $d^2$ collapses to zero (blue horizontal line) signalling that DNA and target are colocalized. Molecule chemical affinity acts as a switch. Here $c = 0.2\%$ and $n_0 = 24$. **Inset** The very same results are obtained by changing the molecule concentration $c$ and the binding sites number $n_0$, as this 3D phase diagram in $(c, E, n_0)$ space shows. The yellow circles mark the transition points between the two phases, the transition surface (grey) is a power-law fit: $c\,(E-\tilde{E})^{\alpha}\,(n_0-\tilde{n_0})^{\beta} = const.$, with $\alpha = 4.5$, $\beta = 1.2$, $\tilde{E} = kT$, $\tilde{n_0} = 8$. **Panel B** The system phase diagram shows the regions in the $(c, E)$ plane where DNA attachment to target and Brownian diffusion occur (here $n_0 = 24$), in the presence (green circles) or absence (orange squares) of non-specific binding sites with a low affinity for molecular binders ($E_{NS} = 1.5kT$).

model: along with the clusters of "specific sites" previously included on the polymer and on its target, we inserted up to $4 \cdot 10^4$ "non-specific" (i.e., low affinity) binding sites distributed on the target surface and within the polymer itself. We performed Monte Carlo simulations to find out the equilibrium status of the system as function of the molecular concentration $c$ and the specific binding energy $E$, with a fixed affinity for non-specific sites equal to $E_{NS} = 1.5kT$.

Fig. 2.12 panel B shows the changes in the phase diagram with respect to the case $E_{NS} = 0$. Orange squares and green circles mark the transition points between the "Brownian" and the "Colocalization" phase respectively for $E_{NS} = 0$ (the case we dealt with previously) and $E_{NS} = 1.5kT$. The plot reveals that the presence of non-specific binding sites moves upward the transition line. This is due to a reduction of the effective concentration of molecules available to the specific sites, that are responsible for recognition and attachment to the target. Such effect can be indeed important and affects the location of the transition line even for comparatively small affinities (e.g., $E_{NS} = 1.5kT$, see Fig. 2.12 panel B). Yet, the overall "shuttling" mechanisms we discussed before is shown to be well robust.

#### 2.5.2.3 Dynamics of "shuttling"

Fig. 2.13 shows the dynamics of $d^2(t)$ for two values of the interaction energy $E$ (here $c = 0.2\%$ and $n_0 = 24$). When the binding energy $E$ is small enough, say $E = 1.6kT$, the long time value of $d^2(t)$ is equal to 100%, as expected for a randomly diffusing polymer (fig. 2.13, red triangles). A drastic change in behavior is observed, however, when the energy is raised to $E = 2.5kT$ (see Fig. 2.13, green diamonds): now the long time plateau of $d^2$ collapses to zero, signalling that a full colocalization has occurred. An effective attraction force is generated and the polymer spontaneously finds and stably binds its target (fig. 2.13 upper panels), as a result of the thermodynamic mechanism described above.

As in the case of DNA sequence pairing (see page 31), the dynamics is char-

acterized at small times by a Brownian diffusion regime where $d^2(t)$ is linear in $t$ (see inset fig. 2.13) and the polymer randomly explores the space around. During that time, it enters in contact with its target. Afterwards, an exponential decay of $d^2(t)$ is observed to the equilibrium value and, for $E$ large enough the interaction is stabilized.

Also in this case, we measured the mean square displacement of the polymer center of mass from the initial position, $\langle \Delta s^2 \rangle(t)$, for the same two values of the interaction energy, $E$, considered above. For $E/kT = 1.6$ (i.e., when no stable colocalization is observed), $\langle \Delta s^2 \rangle(t)$ has the typical Brownian linear behavior with $t$, at short as well as at long times (see fig. 2.14, red triangles): overall, the polymer motion is unaffected by the presence of binding molecules and nuclear scaffold. For $E/kT = 2.5$, instead, Brownian diffusion is only found at short $t$, while the polymer is searching for its target; at longer times, $\langle \Delta s^2 \rangle(t)$ reaches a constant plateau which signals that the polymer becomes firmly bound to the BSs of the nuclear target and cannot diffuse any more (see fig. 2.14, green diamonds).

The inset in fig. 2.14 shows both the short and long time diffusion constants $D_0$ and $D_\infty$ as function of the energy $E$. As expected, the long time diffusion constant, $D_\infty(E)$, has a different behavior, with respect to the case of DNA pairing (see fig. 2.9): when $E$ is small, $D_\infty(E)$ is very close to $D_0(E)$, showing that the polymer motion is diffusive at all times. However, above a transition point, $E_{tr} \simeq 2.1kT$, $D_\infty(E)$ collapses to zero as a result of the attachment of the DNA segment to the fixed scaffold which stops further diffusion (see fig. 2.14 inset and upper panels in fig. 2.13).

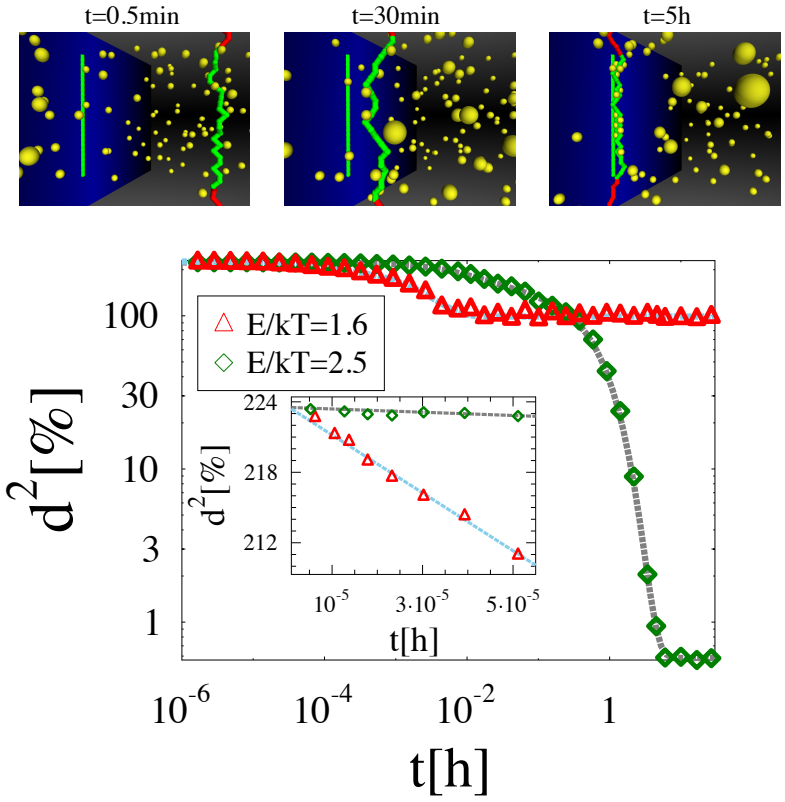


Figure 2.13: The normalized mean square distance, $d^2$, between the DNA and the nuclear target binding sites (BSs) is plotted as a function of time, $t$, for two values of their binding molecule chemical affinity, $E$ (here $c = 0.2\%$ and $n_0 = 24$). Data from Monte Carlo (MC) simulations. For $E = 1.6kT$ (squares), $d^2$ approaches the average distance found for a polymer randomly located in space (i.e., 100% in our normalization). For $E = 2.5kT$ (circles), $d^2$ has a different behavior, as it collapses to zero, showing that shuttling to target has occurred. $d^2(t)$ has two time regimes: a linear behavior at small $t$ (see **inset**) due to the initial Brownian diffusion ("target searching" regime) and a long time exponential approach to the equilibrium value ("target recognition" regime). The latter corresponds to colocalization only if $E$ is above a threshold (see fig. 2.12). The **upper panels** show system configurations from MC simulations at three time periods for $E = 2.5kT$, and provide a pictorial representation of our model: the DNA locus is modelled as a self-avoiding walk (SAW) polymer made by "beads" which have an affinity equal to 0 or to $E$ (red beads and green beads respectively) for Brownianly diffusing molecules (yellow beads). A cluster of binding sites is also present on the nuclear target (blue surface).

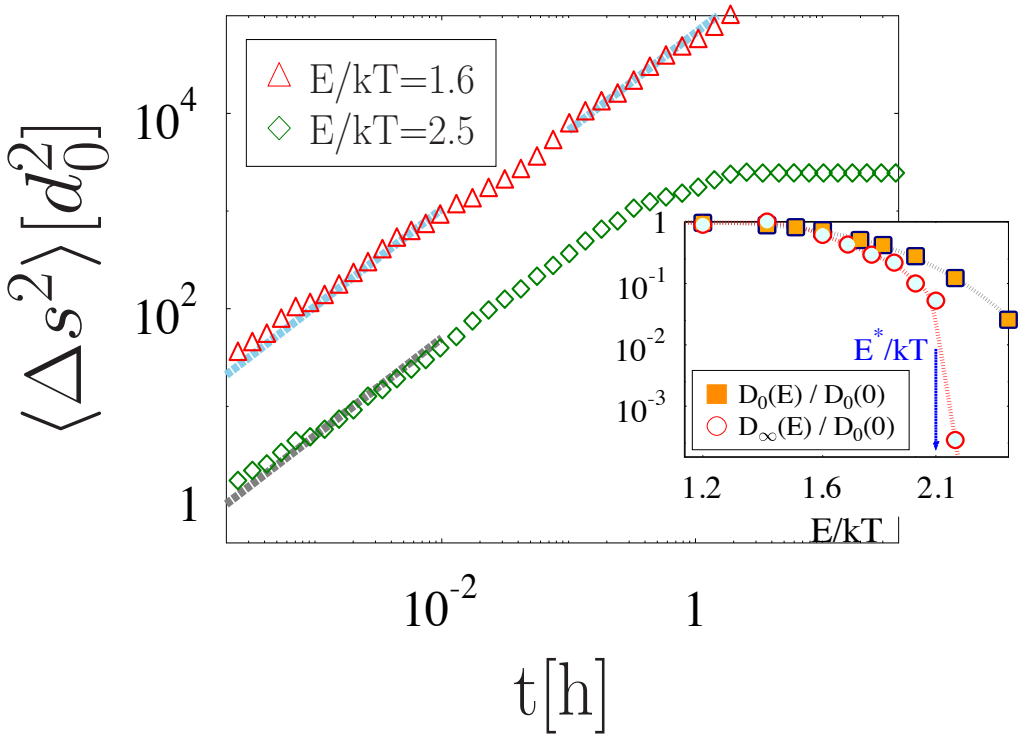


Figure 2.14: **Main panel** The mean square displacement, $\langle \Delta s^2 \rangle$, of the center of mass of the DNA polymer is plotted as function of time for two values of $E$. For $E/kT = 1.6$ (squares) $\langle \Delta s^2 \rangle$ is linear in $t$ at all times, signalling a purely Brownian motion: a linear fit at short and long times (green lines) returns approximately the same diffusion constant ($D_0$ and $D_\infty$, see Inset). For $E/kT = 2.5$ (circles), when colocalization occurs, after an initial Brownian regime (grey linear fit), around $t = 2h$, $\langle \Delta s^2 \rangle(t)$ reaches a plateau, showing that a stable contact with target is established and diffusion interrupted. **Inset** The short and long time diffusion constants, $D_0$ and $D_\infty$, are shown as a function of $E$ (normalized by $D_0(E = 0)$). $D_\infty$ is very close to $D_0$ only for $E < E^* = 2.1kT$ (blue arrow), else it goes to zero indicating that long time diffusion is halted.

## 2.6 Conclusions

In the cell nucleus, in a striking example of self-organization, the architecture of a vast number of DNA and nuclear loci is orchestrated to form complex and functional patterns involving regulatory contacts. In most cases active processes are not required for colocalization [54, 19, 21] and questions arise on how DNA sequences recognize their targets (other DNA sequences or nuclear structures) and establish their relative positioning, and how the cell can control these processes. The pairing of X chromosomes at the onset of XCI as well as the "shuttling" of X's at the end of XCI process (see previous chapter), are some prominent examples of these phenomena.
Starting from the experimental data, via a schematic Statistical Mechanics model, here we tried to address these questions and to propose a first quantitative scenario of a colocalization mechanism based on weak, biochemically unstable interactions between specific DNA sequences and their molecular binders. The mere production of molecules which bind both DNA and target is not sufficient to produce reliable and stable contacts. We showed they are activated only above a phase transition point, i.e., for concentration/affinity of the molecular mediators above precise threshold values (e.g., molecule concentrations around $\rho \sim 10^{-3} \div 10^{-1} \mu$mol/litre correspond to transition energies in the range $E \sim 3 \div 7kT$). Once these conditions are met, DNA loci find their relative positions as stable thermodynamic states at no energetic costs, as the resources required are provided by the surrounding thermal bath.

The switch-like nature of the mechanism of target recognition and colocalization we discussed could be exploited in the cell to reliably induce loci colocalization. In fact, well known cell strategies of modification of chromatin structure (i.e., change $E$ or $n_0$) or upregulation of binding proteins (i.e., change $c$) can produce precise, switch-like, architectural rearrangements.

Deep similarities are found across a variety of experimental data like those discussed in the section 2.2, including specific aspects such as the effects of protein concentration changes on DNA looping (see, e.g., [34, 38]).

The robust thermodynamic essence of the process we discuss could support the idea that "passive shuttling" phenomena can be traced back to simple universal mechanisms (see, e.g., [81, 17, 35, 20, 78]), in a sense independent on the biochemical details found in specific cases. On the other hand, many complexities can arise in real cell nuclei where a variety of other specific mechanisms are likely to intervene.

For sake of definiteness, we referred to DNA, but similar thermodynamic mechanisms could work for other biological polymers, such as RNA, etc. Non-specific molecular factors and unspecific DNA binding can further help the search kinetics [8, 90, 89], while other processes can intervene, e.g., to stabilize binding and to adjust DNA/target alignment if necessary. We also showed that non specific DNA/target binding sites can have an important effect on colocalization, yet the general scenario depicted above is unchanged. Testable predictions about the outcomes of, e.g., genetic/chemical manipulations (such as DNA deletions) can be made, which can be tested against experimental data.

We tried to set the system parameters (e.g., the molecule concentration, the dynamics time scale) in a regime relevant to the real biological cases (see section 2.3). Nevertheless, our model is very schematic and we included only the minimal molecular ingredients (say, molecular binders and specific DNA sites) which emerge from the experiments. Yet, a simple model could better serve the purpose to illustrate the core ingredients necessary to DNA-target recognition (which can be traced back to polymer adsorption) and to depict a schematic but quantitative scenario.

# Chapter 3

# X random choice via a Symmetry Breaking mechanism

We present here a possible mechanism, based on a thermodynamic phase transition, which the cell could exploit to randomly choose the inactive X and which also accounts for the changes in the DNA architecture and the X chromosome pairing observed experimentally (see section 1.4) [77, 72].

## 3.1 Introduction

We described in a previous section (1.4), the recent experimental insights into the complex and dynamic regulation of the XCI key control region called "X Inactivation Center" (*Xic*), and, in particular, its different spatial organization on the two X chromosomes. We discuss here a model to explain, on quantitative grounds, how the *Xic* locus is orchestrated through such a complex pattern of conformational changes; how its elements recognize each other and colocalize; and how the random, yet mutually exclusive choice of fate is later determined, i.e., how the intrinsic symmetry of the X's is spontaneously broken.
Starting from the above experimental information, our model (see Fig.3.1) predicts the existence of two molecular regulators, say, type-A and B molecules.

It poses that along the DNA locus there are two regions (type-a regions), in the area around *Jpx* and *Xist*, having an array of binding sites for type-A molecules, which can bridge them to another key control region (type-c region) expected to be located in the area of the "buffer". Analogously, there are two type-b regions, around *Tsix* and *Xite*, where type-B molecules can bind and form a bridge with the same type-c region above.
We show on physical grounds that the above minimal elements are sufficient to orchestrate conformational changes and symmetry breaking as seen at XCI, and we explain how they spontaneously induce the complex regulation of the locus. In particular, we show (see Fig.3.1) that if molecule concentration/DNA-binding energy rises above threshold, the distinct molecular factors produce long distance interaction (recognition) and stable colocalization of their corresponding DNA binding sites, thus forming a type-a and a type-b "hub". This is symmetrically achieved on each of the X chromosomes (see fig. 3.1 panel B). The symmetry is spontaneously broken when a mutual interaction is turned on between molecules of the same kind. We show that molecule cooperativity produces the self-assembling of a single major aggregate for each class which binds randomly, yet in a mutually exclusive way, only one X. This, in turn, leads to two different architectures on the X's, as the X bound to type-A molecular aggregate will show only a type-a hub, whereas a type-b hub will be found on the other X bound by the type-B aggregate (see fig. 3.1 panel C).

The colocalization and symmetry breaking mechanisms derive from two distinct thermodynamic phase transitions [11]. Their switch-like nature can explain how a sharp regulation of nuclear architecture and stochastic choice of fate can be obtained by simple strategies, such as protein upregulation or chromatin modification. Importantly, the model predicts energy/concentration thresholds which are in the expected biological range (weak biochemical energies, fractions of micro-mole/litre concentrations). It may help understanding the deep connection between architectural changes and regulation of XCI, providing a unified description of such a complex process.

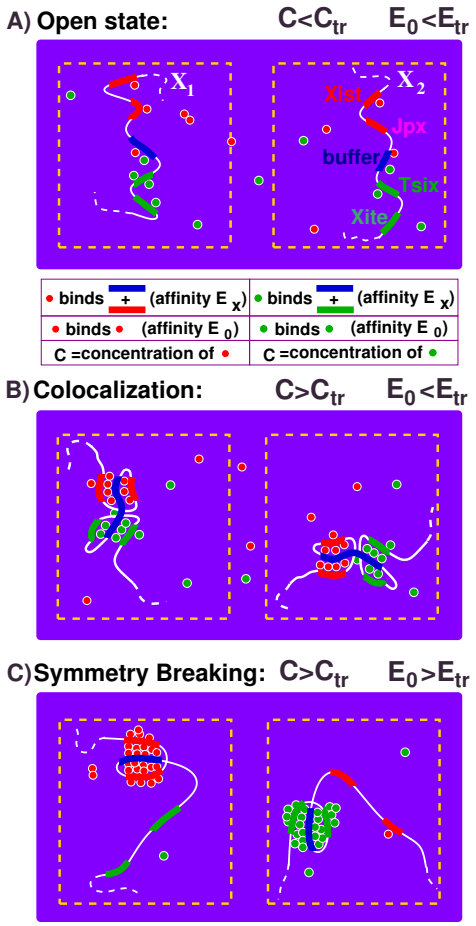


Figure 3.1: A pictorial description of our model. **A)** Two polymers are considered, each having two type-a (red), two type-b (green) and a type-c (blue) regions. Type-a and type-c can be bridged by type-A molecules (red circles) with an affinity $E_X$; type-b and type-c by type-B molecules (green circles). Each molecular species has a concentration $c$. Type-A molecules have also mutual interactions of affinity $E_0$, and similarly type-B ones. The presumptive mapping areas on the Xic are also illustrated (in the upper right panel). There are threshold values for $c$ and $E_0$ (for a given $E_X$), $c_{tr}$ and $E_{tr}$, where the system collective behaviour spontaneously changes. For $c < c_{tr}$ and $E_0 < E_{tr}$, the polymers are statistically found in an open and randomly folded state (as in panel A). **B)** For $c > c_{tr}$, type-a and type-b regions colocalize with type-c and a two hub conformation is established. **C)** If also $E_0 > E_{tr}$, a symmetry breaking occurs as the type-a hub persists on one, randomly selected, polymer (where type-b hub is released), whereas the other polymer takes the opposite conformation. This results from the self-assembling (for $E_0 > E_{tr}$) of a single major cluster of type-A and type-B molecules which compete for the binding sites of type-c region.

## 3.2 Model parameters and Monte Carlo computer simulations

We represent the relevant region on each X chromosome (see Fig. 3.1) by a standard model of polymer physics, a self-avoiding bead chain [23]. As mentioned above, we assume that along each polymer there are two type-a, two type-b and one type-c region. Each polymer bead in these regions can bind up to four molecules of the same type: beads in type-a (respectively type-b) regions can bind only to type-A (respectively type-B) molecules, whereas beads in type-c regions can bind either to type-A or type-B molecules. Moreover, type-A molecules (respectively type-B) can bind, at most, to one type-a (respectively type-b) and one type-c binding site at a time. For simplicity, with no loss of generality of our results, we consider the symmetric case where both the types of molecules have a concentration $c$ and their affinity to DNA is equal to $E_X$ for all the types of loci they can bind to. Similarly, if $n_0$ is the number of type-c binding sites on one chromosome, we assume that each of the type-a and type-b regions contain the same number of binding sites equal to $n_0/4$. The value of $n_0$ is chosen to be of the order of the binding sites number found experimentally in the locus for Ctcf protein (see section 1.4; here $n_0 = 20$).

Type-A (resp. type-B) molecules can bind, with multiple valency, each other with an affinity $E_{AA}$ (resp. $E_{BB}$). For simplicity, we set $E_{AA} = E_{BB} \equiv E_0$ and, considering the number of binding domains available on a Ctcf molecule, the valency of four.

Summarizing, we consider a physical model which includes two identical self-avoiding polymers with two kind of binding sites, for molecules A and B. The polymers have a key control region (type-c) where either A or B can bind. Each kind of molecule has a concentration $c$ and an affinity, $E_X$, for its corresponding binding sites on the polymer. Molecules of type A can also bind each other, with $E_0$, and similarly the B molecules.

The system is investigated by Metropolis Monte Carlo (MC) simulations

[10]. For computational purposes, the system lives on a cubic lattice with lattice spacing, $d_0$. The value of $d_0$ corresponds to the typical size of a DNA binding site and can be roughly estimated to be about three orders of magnitude smaller than the nucleus diameter, say, $d_0 \sim 10nm$ (i.e., a DNA sequence of about $\sim 30bp$). To reduce computation time, we do not simulate the whole nuclear space, but only consider a cubic lattice region with periodic boundary conditions, around each polymer of linear size $L$ (in units of $d_0$); yet, type-A and B molecules are allowed to diffuse from one to the other region. Below, we use lattices with $L = 32$ and periodic boundary conditions. Polymer chains have $n = 96$ beads, but we explored the range $L, n = 32 \div 128$ to check the robustness of our results. Averages run over up to $10^3$ simulations from different initial configurations.

As we explained in the previous chapter (see section 2.3), the fraction, $c$, of molecules per lattice site is related to the molar concentration, $\rho \sim c/d_0^3\mathcal{N}_A$, where $\mathcal{N}_A$ is the Avogadro number. A typical nuclear protein concentration $\rho \sim 1\mu$mole/litre would correspond to $c \sim 0.1\%$: below we consider the range $c \sim 10^{-3} \div 10^0\,\%$.

Diffusing particles (molecules and polymer beads) randomly move from one to a nearest neighbor vertex on the lattice, where no more than one particle can be present at a given time (single occupancy). Chemical interactions are only permitted between nearest neighbour particles.

In our Monte Carlo simulations, the probability of a particle to move to a neighbouring empty site is proportional to the Arrhenius factor $r_0 \exp(-\Delta E/kT)$, where $\Delta E$ is the energy change in the move, $k$ the Boltzmann constant and $T$ the temperature. The prefactor $r_0$ is the bare kinetic rate and gives the scale to convert Monte Carlo to real time: the Monte Carlo time unit, $\tau_0 = r_0^{-1}$, is the time to try to update once, on average, all the particles of the system [10].

We set $r_0$ here by imposing that the polymer diffusion coefficient, $D$, has the same order of magnitude than real mammalian DNA loci (as we did above, section 2.3). We exploit the defining relation: in three-dimension

$D = \langle \Delta s^2 \rangle \, (d_0^2/6\tau_0)$, where $\langle \Delta s^2 \rangle$ is the mean square displacement (expressed in units of $d_0$) of the polymer center-of-mass per unit MC time. In our system we find values around $\langle \Delta s^2 \rangle = 1.5 \cdot 10^{-3} \, d_0^2$. The order of magnitude of measured diffusion constants of, e.g., human DNA loci is $D \sim 1 \mu m^2/hour$ [14], thus we obtain $\tau_0 \sim 90 \mu s$, a value falling within the range of known biochemical kinetic constants [88].

To monitor the spatial configuration of the polymers, we measured several quantities, such as the distance between the type-a and type-c regions defined as:

$$d = \frac{1}{n_0/4} \sum_{i=1}^{n_0/4} min \left\{ d_{a_i,c_j} \right\}_{j=1}^{n_0} \tag{3.1}$$

where $d_{a_i,c_j}$ is the distance between the $i$-th bead of the type-a region and the $j$-th bead of type-c region. For each type-a bead, the minimum of these distances is calculated, and then averaged over all the $n_0/4$ beads of the type-a region. Of course, the same definition is used for the distance between type-b and type-c regions. A type-a (resp. type-b) region is defined to be "colocalized" with the type-c region if $d < 2d_0$.

## 3.3 Pre-XCI X conformation

We first show that Brownian diffusing molecules can produce a symmetric polymer conformation where two "hubs" are formed by the colocalization of type-a and of type-b with type-c region. The mechanism acting here is the DNA "passive-shuttling" which we described in details in the chapter 2: it is based on a thermodynamic process (a phase transition) which acts switch-like when concentration/affinity of binding molecules rise above a threshold.

For sake of simplicity, we consider first the case where molecule mutual interaction is turned off, $E_0 = 0$, and assume that initially the two polymers have a randomly open conformation, as found in nuclei before XCI [85].

A type-A molecule (red circles in Fig. 3.1) can form a bridge between type-a and type-c regions, an event resulting from the stochastic double encounter

of the molecule with its targets. If molecule concentration, $c$ (or $E_X$, see below), is small the half-life of such a bridge is short and the regions on average float away from each other: their contact is totally unstable. At higher $c$ (or $E_X$), however, a positive feedback mechanism occurs: the higher the number of bound molecules to type-a and type-c regions, the larger the number of bridges which reinforce each other and stabilize the conformation, as multiple bonds should be simultaneously broken to release the contact. The argument predicts, thus, a threshold concentration where such a positive feedback mechanism becomes winning and a stable contact is established between type-a and type-c regions.

This pictorial scenario is confirmed by Monte Carlo simulations. We measured the colocalization order parameter, $\pi = (p_A + p_B)/2$, where $p_A$ (respectively $p_B$) is the probability of colocalization of type-a regions (respectively type-b). More precisely, $p_{A/B}$ is the fraction of cases where the region mean distance, at equilibrium, is less than $2d_0$. The order parameter, $\pi = \pi(c, E_X, E_0)$ is close to zero, $\pi \simeq 0$, if neither type-a nor type-b regions are in contact; in case only one pair is stably colocalized then $\pi \simeq 0.5$; finally, $\pi \simeq 1$ if both hubs are formed.

Figure 3.2 panel A shows the time evolution of $\pi(t)$ in two cases (from an open initial configuration, here $E_X = 3kT$): if $c$ is small, $\pi$ remains indefinitely close to zero, $\pi \sim 0$, as no stable contact is statistically possible; instead, if $c$ is high enough, $\pi$ grows to a value close to one, $\pi \simeq 1$, showing that both type-a and type-b hubs are formed.

In the space of the control parameters, $(c, E_X)$, a sharp line separates the two regimes, as shown in Fig. 3.2 panel B: when $c$ or $E_X$ are small, contacts cannot be stable; conversely, above the transition line the two hubs conformation is reliably established on each polymer. Such a line marks the boundary between two thermodynamic phases [11]: it corresponds to the point where the entropy loss due to hub formation is compensated by the energy gain obtained from the establishment of the corresponding bridges. For its thermodynamics nature the process is strongly robust (see also chapter

2).

The discovery of such a switch-like behaviour can also explain how hub formation can be sharply and reliably regulated in the cell by increasing the concentration of specific molecular mediators or the affinity to their DNA target sites, e.g., by chromatin or molecule modifications.

The position of the transition line is also dependent on the number of available binding sites, $n_0$, since, schematically, the overall binding energy scale is $n_0 E_X$. We predict, thus, the presence of non-linear threshold effects in genetic deletion/insertions of the locus (see section 2.5.1.3).

From Monte Carlo results we can predict concentration (or energy) thresholds in real nuclei. Consider the case where $E_X$ is of the order of a typical Transcription Factor binding energy, i.e., a few units in $k_B T$ [88]. In that case our Monte Carlo predicts thresholds around $c \sim 0.1\%$, i.e., molar thresholds around $\rho \sim 1\mu$mole/litre, a value in the range of typical nuclear protein concentrations. Higher $E_X$ would correspond to lower $\rho$ (see Fig. 3.2 panel B). For instance, "in vitro" measures of Ctcf DNA binding energies give $E_X \sim 20kT$ (see, e.g., [70, 71]), predicting molar concentration thresholds in the sub-fraction of $\mu$mole/litre.

Finally, the mechanism leading to stable hub formation has to be fast enough to serve functional purposes. Real dynamics is only very schematically captured by Monte Carlo Metropolis simulations, which are considered though to well describe the general long time evolution of a system dominated by Brownian processes. In our model we find colocalization times of the order of minutes as shown in Fig. 3.2 panel A, a scale consistent with biological expectations.

## 3.4 Stochastic choice of fate

The hub formation mechanism illustrated above is fully "symmetric" on the two X polymers. Now we show that turning on molecule reciprocal interaction, $E_0$, is enough to trigger a different thermodynamic mechanism which

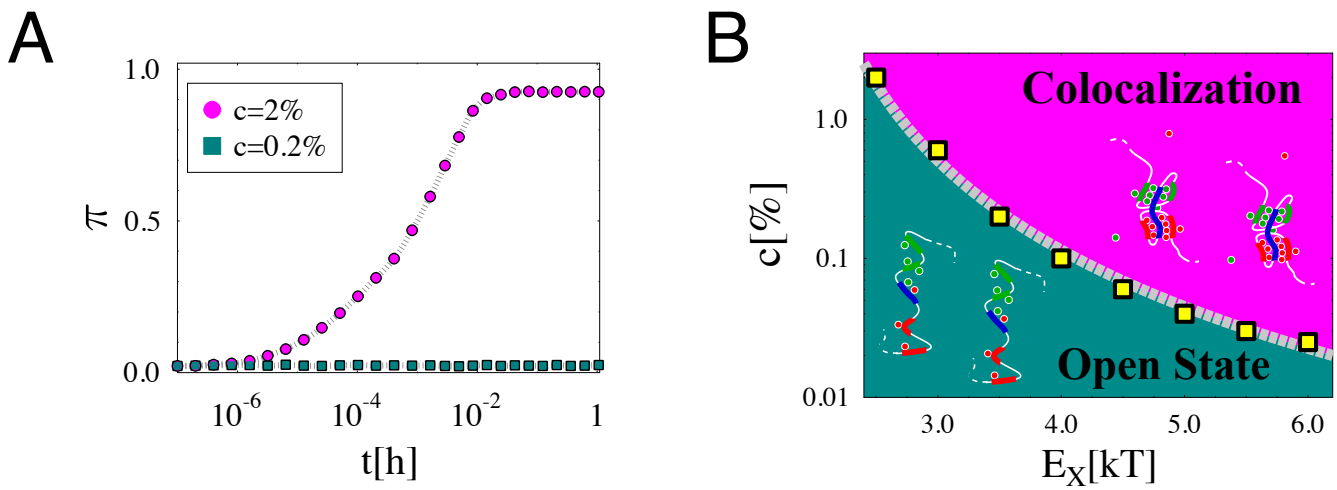


Figure 3.2: **Panel A** The colocalization order parameter, $\pi = (p_A + p_B)/2$, is the average probability of type-a and type-b regions to colocalize with type-c region. Its time evolution, from an initial open polymer conformation, is shown for two values of concentration, $c$ (here $E_X = 3kT$ and $E_0 = 0$). For $c = 0.2\% < c_{tr}$, $\pi$ is zero at all times: neither type-a nor type-b regions succeed in forming stable contacts with type-c. For $c = 2\% > c_{tr}$, after a transient of the orders of minutes, $\pi$ approaches one: full colocalization is established. **Panel B** The colocalization phase diagram in the $(E_X, c)$ plane (for $E_0 = 0$): in the region below the sharp transition line, $c_{tr}(E_X)$ (gray dashed line), the polymers are found in an open state; above $c_{tr}(E_X)$, they exhibit the colocalization of type-a and type-b with type-c region. The grey dashed line is a power-law fit: $c = (E_X - \tilde{E}_X)^{-\alpha}$ with $\tilde{E}_X = 1.7kT$ and $\alpha = 2.6$.

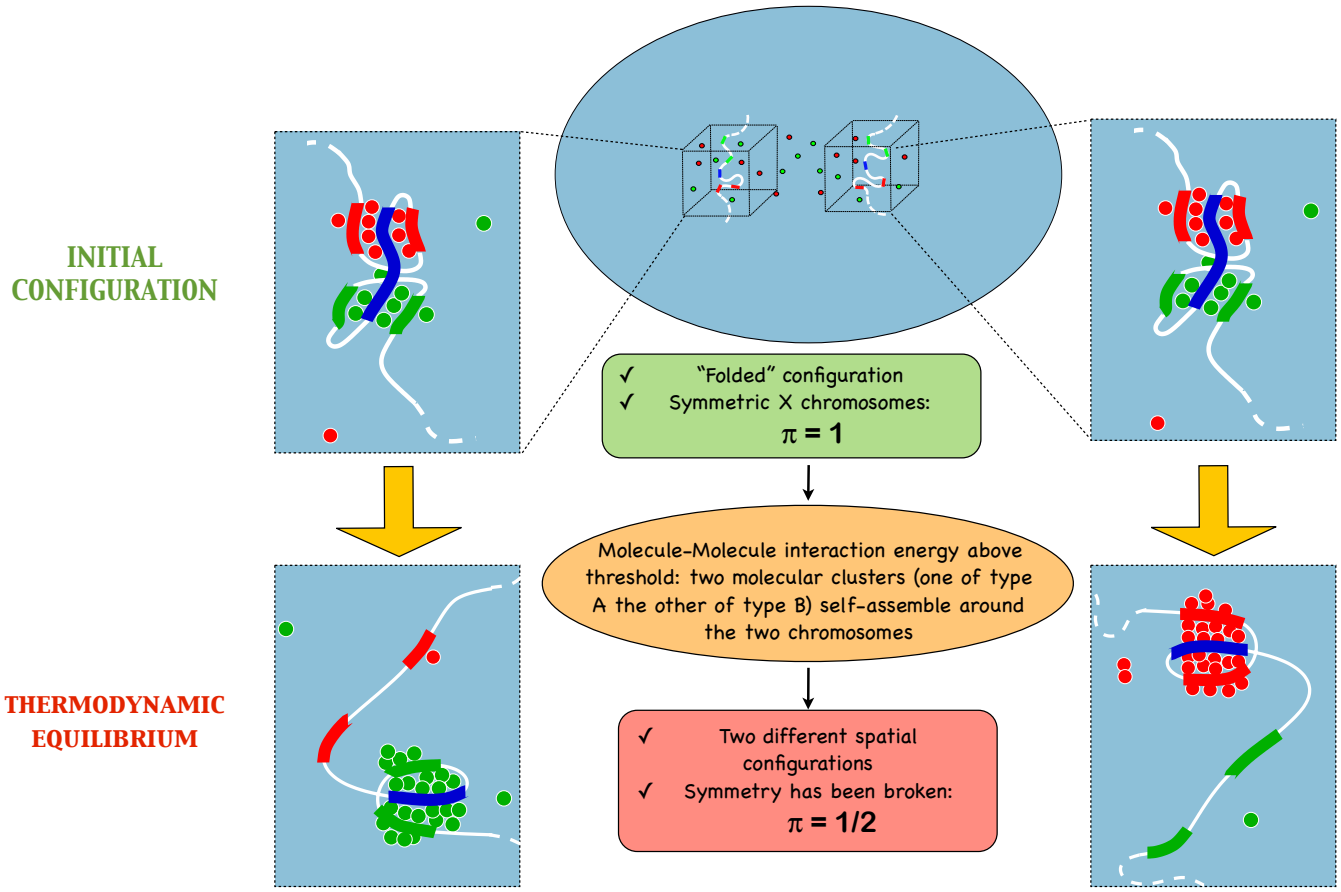


Figure 3.3: A schematic representation of the symmetry breaking mechanism is illustrated. We started from symmetric, "folded" configurations, in which type-a (resp. type-b) regions are colocalized with type-c with a probability $p_A$ (resp. $p_B$) equal to 1. So, the pairing order parameter $\pi = (p_A + p_B)/2 = 1$. When an above threshold interaction energy between the two types of molecules is turned on, two clusters are assembled around the two chromosomes and the symmetry is broken ($\pi = 1/2$).

produces the breaking of the X symmetry: a single major type A molecule aggregate is formed which randomly binds one X (where the B linked hub will thus open); analogously, a major type B molecule aggregate binds only the other X (where the type-a hub will open). This results from molecule binding cooperativity: if $E_0$ (and $c$, see below) is above a critical threshold, the energy gain in forming a single major cluster of molecules of the same kind (that maximizes the number of chemical bonds between molecules) compensates the corresponding entropy reduction. The single aggregate will be randomly bound around just one X, leaving the other X "naked" of that kind of molecule.

Type-A and B major clusters bind opposite X polymers because A and B molecules compete for the binding sites in the type-c region. Hence, if on one X a fluctuation increases the presence of, say, A molecules, cooperativity tends to favour their assembling at that site and B molecules are expelled (a state with an equal number of A and B would be unstable); in turn, the depletion of A around the other X favours the assembling of B molecules on it. On the X where the A-cluster binds the key regulator type-c region, the B-related loci can no longer be stably linked, and their hub opens; the opposite situation happens on the other X (see fig. 3.3).

The above scenario is illustrated by Monte Carlo simulations where the symmetry breaking order parameter, $m_A = |\rho_A^{(1)} - \rho_A^{(2)}|/(\rho_A^{(1)} + \rho_A^{(2)})$, is measured (here $\rho_A^{(i)}$ is the average local concentration of $A$ molecules around type-c region of polymer $i = 1, 2$): $m_A$ is close to zero if an equal amount of A molecules is present around the two X polymers, whereas it approaches one if the symmetry is spontaneously broken ($m_B$ behaves analogously). Figure 3.4 shows the time evolution of $m_A(t)$ from an initial configuration corresponding to the symmetric state where each X polymer has two hubs. It reports two cases: if $E_0$ is small, $m_A$ remains close to zero at all times; conversely, if $E_0$ is high enough, $m_A$ approaches one because A molecules reside mostly around just one, randomly chosen polymer and the symmetry is broken.

To illustrate further the process, figure 3.5 panel A shows the corresponding dynamics of $\rho_A^{(1)}$ and $\rho_A^{(2)}$: in the example pictured, $\rho_A^{(1)}$ strongly increases, whereas $\rho_A^{(2)}$ plunges approximately to zero. There is an initial transient when $\rho_A^{(1)}$ and $\rho_A^{(2)}$ behave similarly, yet $\rho_A^{(1)}$ prevails as soon as a fluctuation favouring the A's around the chromosome 1 becomes spontaneously amplified. Thus, type-c region on polymer 2 is depleted of A molecules and the hub they hold is released, i.e., the type-A regions contact opens as also illustrated in fig. 3.5 panel B, where $p_A$ on chromosome 1 and 2 are reported. Of course, the same quantities plotted for type-B molecules and type-b regions, would show exactly the opposite situation, with the type-b hub released on chromosome 1.

The Symmetry Breaking mechanism is also switch-like, as the phase diagram of figure 3.6 proves: in the $(c, E_0)$ space, as soon as a narrow transition line is crossed the system switches from an X symmetric to a broken X symmetry state. As far as XCI is concerned, the predicted single B molecule aggregate is interpreted as a $Xist$ repressing factor (a Blocking Factor, BF) and designates the future active X. The A aggregate marks the X where $Xist$ is enhanced by $Jpx$ and is interpreted as an activating factor. Importantly, the thresholds predicted by our theory for the symmetry breaking mechanisms also fall is the correct biochemical range (see above and Fig. 3.6).

The time scale required to break the symmetry in a real nucleus can depend on a number of details, such as its specific spatial arrangement or the nature of the chemical reactions involved. Our Monte Carlo can, thus, provide only a very rough order of magnitude estimate. As shown in Fig. 3.5, such a time scale is predicted to be around 10 hours, a value not too far from the biological time scale required for XCI initiation [93].

## 3.5 Pairing and Random Choice

We saw in the previous chapters that the two X chromosomes have been shown to pair off during the XCI process.

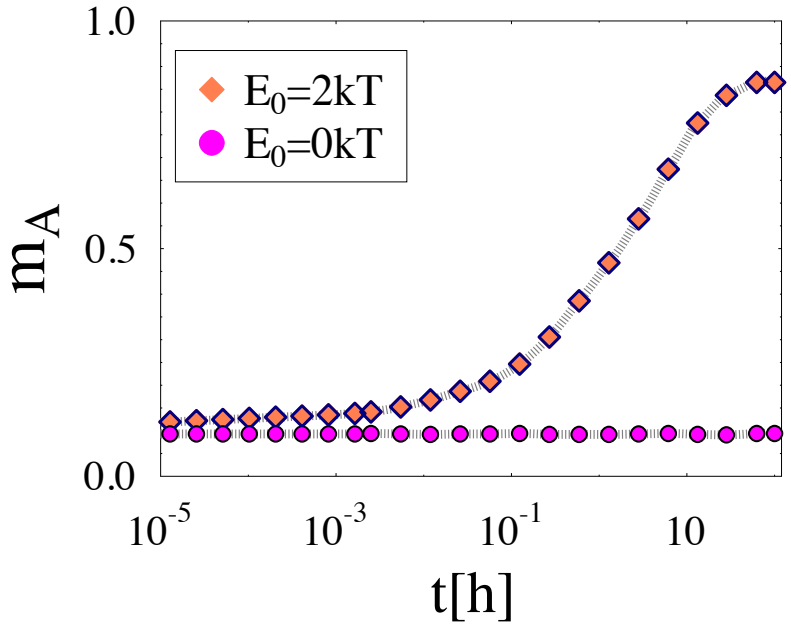


Figure 3.4: The symmetry breaking order parameter, $m_A = |\rho_A^{(1)} - \rho_A^{(2)}|/(\rho_A^{(1)} + \rho_A^{(2)})$, is the average difference of type-A molecule density around type-c region of polymers 1 and 2, divided by their sum. Its dynamics, from a "colocalized state", is shown for two values of $E_0$ (here $c = 2\%$ and $E_X = 3kT$). If $E_0 = 0 < E_{tr}$ (purple circles), $m_A$ is close to zero, i.e., $\rho_A^{(1)} \simeq \rho_A^{(2)}$: molecules are equally distributed around polymers. If $E_0 = 2kT > E_{tr}$ (orange diamonds), after a transient of about ten hours, $m_A$ rises close to one, i.e., either $\rho_A^{(1)} \to 0$ or $\rho_A^{(2)} \to 0$: molecules have aggregated around only one of the polymers whose binding symmetry is broken.

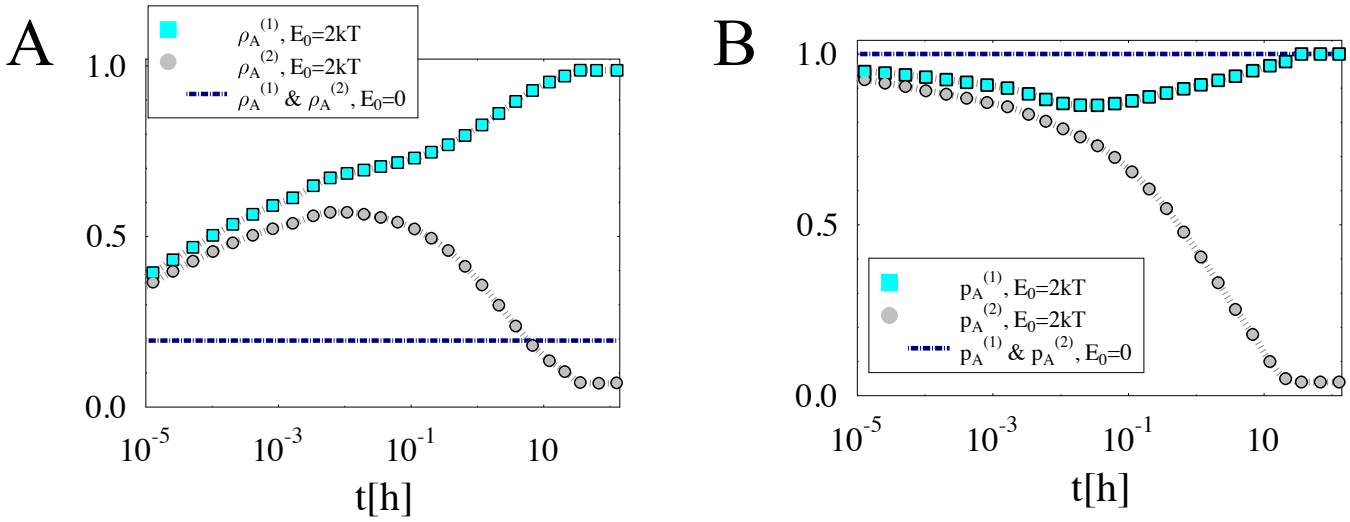


Figure 3.5: An example of the dynamics of $\rho_A^{(1)}(t)$ and $\rho_A^{(2)}(t)$, i.e., the type-A molecule density around type-c region of polymers 1 and 2 (**panel A**), and of the colocalization probabilities $p_A^{(1)}(t)$ and $p_A^{(2)}(t)$ (**panel B**) of type-a regions on polymer 1 and 2. In this example, with $E_0 = 2kT$ (here $c = 0.2\%$ and $E_X = 3kT$), $\rho_A^{(1)}(t)$ grows to one while $\rho_A^{(2)}(t)$ plunges to zero, and at the same time $p_A^{(1)}(t) \to 1$ whereas $p_A^{(2)}(t) \to 0$. The dashed blue lines indicate the values of $\rho_A^{(1)} = \rho_A^{(2)}$ and $p_A^{(1)} = p_A^{(2)}$ when $E_0 = 0$

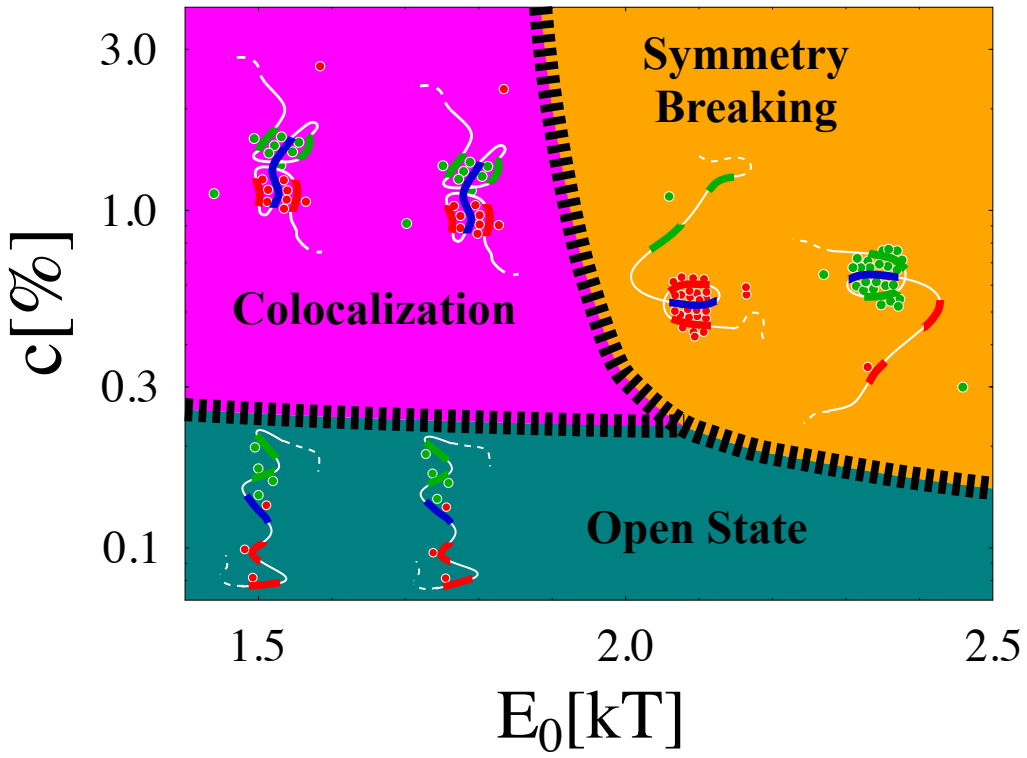


Figure 3.6: The system phase diagram in the $(E_0, c)$ plane (for $E_X = 3kT$) has three phases. If $E_0$ is below the transition line, $E_{tr}(c)$, the system is in one of its symmetric phases: the "open state" phase (at low $c$) or the "colocalization phase" (at high $c$). If $E_0 > E_{tr}(c)$, the polymers conformation symmetry is broken: the type-a hub persists only on one, randomly chosen polymer, whereas the type-b hub persists on the other.

Our Symmetry Breaking model also predicts that the X chromosomes colocalize (precisely at the $Xic$ region) on a longer time scale, as found in recent experiments [5, 93]. When the "symmetry breaking" molecular aggregates have bound their target, the homologous regions on the opposite X (which is "open") can eventually diffuse (with the rest of its X) to be bound itself; that process occurs on a much longer time scale because it involves diffusion of whole DNA regions rather than molecules. Yet, symmetry breaking would be followed by a form of X colocalization unless other phenomena intervene. We tested such a scenario with the focus on just one of the DNA regions (say, only type-a), to speed-up simulations. And, in that context, we also explored the possibility that many a molecular factor is involved in the symmetry breaking process.
So far, to illustrate the key concepts of colocalization (see chapter 2) and symmetry breaking, we employed a realization of the model which includes very minimal ingredients. In fact, a number of variants can be considered to account for the variety of complications likely to occur in real nuclei. First of all, the number of molecular factors involved in the process can be much larger than just two (e.g., the molecule mutual interaction could be turned on by other factors). Specific factors, e.g., non coding local transcripts [46], and unspecific chromatin proteins, such as Ctcf (or Yy1 and Oct4 [25, 24]), are likely to cooperate in the timeline of such a complex process.

We considered two directed self-avoiding bead chain polymers which can bind a concentration of diffusing molecules (type-1), with energy $E_X$. These molecules can form a bridge between the beads of the different polymers and also bind (energy $E_{12}$), at most, two molecules of a different species (type-2), which, in turn, have multiple valency mutual interactions (energy $E_{22}$). We used a lattice with periodic boundary conditions having dimensions $L_x = 2L$, $L_y = L$ and $L_z = L$ in units of $d_0$, the lattice space constant. In the simulations here discussed $L = 16$, but we checked the results with $L$ in the range $16 \div 64$. The polymers have $n = L$ beads, which act as binding sites for type-1 molecules (see fig. 3.7). We run Monte Carlo computer simulations of

this system, by using the Metropolis algorithm (see section 3.2 for details). We take into account the faster dynamics of the molecules with respect to the DNA loci by fixing a molecule diffusion constant $5 \cdot 10^3$ times larger than that of the DNA loci, as experiments suggest (see [68] for typical diffusion constants of nuclear proteins and [14] for DNA loci). We measure the average distance between the polymers defined as:

$$d = \frac{1}{n} \sum_{i=1}^{n} \langle r(z_i) \rangle \tag{3.2}$$

$\langle r(z_i) \rangle$ being the distance between the polymer bead at height $z_i$, averaged over different simulations and over all the $n$ polymer beads. The polymers are considered to be "colocalized" if their distance is less than $3d_0$.

In such a system, we found that, if the concentrations/binding energies of the molecules are high enough, type-1/type-2 molecular complexes form a unique cluster around one of the two polymers, so breaking the binding symmetry of the polymers. Then, the couple of polymers eventually colocalize at two orders of magnitude longer time scales.

We show this in fig. 3.8: panel A illustrates the pairing probability $p$ of the polymer's couple (blue circles) and the order parameter $m_2$ (orange squares) defined as $\left(|\rho_2^{(1)} - \rho_2^{(2)}|\right) / \left(\rho_2^{(1)} + \rho_2^{(2)}\right)$, where $\rho_2^{(i)}$, $i = \{1, 2\}$, is the density of type-2 molecules around the $i-$th polymer ($m_1$ has the same behaviour). In panel B a 2D projection of the lattice is shown at three time frames: the green spots are the polymers centre-of-mass, whereas the density of type-2 molecules is represented in colour scale (blue areas correspond to low densities, red to high densities). In this simplified version of the model, we fixed the time scale by imposing that the time required to break the symmetry is the same we found in the previous model (see section 3.4).

Initially the polymers are placed at the maximum possible distance and the molecules are randomly distributed in the lattice (upper panel B). It is seen that $m(t) \sim 80\%$ at $t \sim 0.1 \div 1h$, indicating that type-2 molecules have clusterized around one polymer (middle panel B). Yet, with a slower dynamics, polymers continue to diffuse, and eventually colocalize at $t \gtrsim 10h$: this is

signalled by the raise of $p$ to $\sim 100\%$, and the corresponding falling of $m$ to $\sim 0$, as the molecular cluster now includes both the polymers (lower panel B).

In the symmetry breaking model we illustrate in the previous section, X colocalization can be, of course, prevented by other intervening nuclear events, or more trivially by the specific chemical details of the system. Suppose, for instance, that type-A and B molecules have a structure whereby DNA can be bound only if they are not previously bound to other type-A and B molecules: in this case, after the A and B aggregate has bound its target, the eventually approaching homologous region of the other X cannot be bound, and stable colocalization of the X would not occur.
Most of these details are still mysterious and our analysis could help individuating candidate elements and designing targeted experiments.

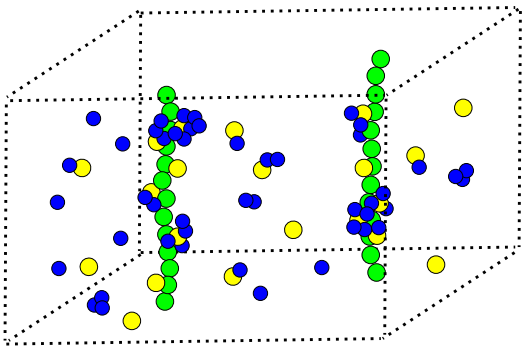

Figure 3.7: Pictorial representation of the model system we used to investigate the relationships between the symmetry breaking mechanism and the pairing. In a lattice with periodic boundary conditions, two directed polymers are included having binding sites (green beads) for a set of diffusing molecules (type-1 molecules, yellow spheres). These molecules can form bonds with another type of molecules (type-2, blue spheres), which, in turn have a reciprocal interaction energy.

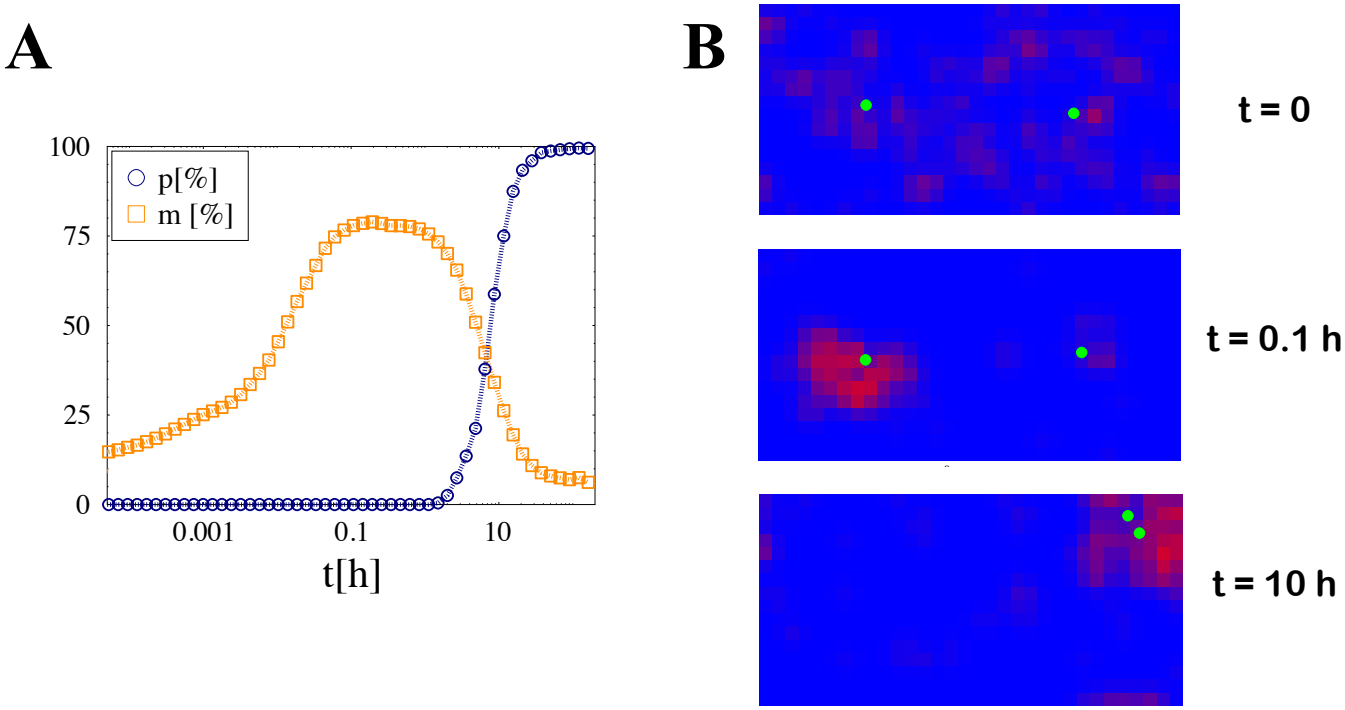


Figure 3.8: **Panel A** The pairing probability $p$ of the two polymers (blue circles) and the order parameter $m_2 = \left(|\rho_2^{(1)} - \rho_2^{(2)}|\right) / \left(\rho_2^{(1)} + \rho_2^{(2)}\right)$ (orange squares; $\rho_2^{(i)}$ is the density of type-2 molecules around the $i$-th chromosome) are plotted as function of time. It is shown that the molecules break the binding symmetry between the polymers ($m \sim 80\%$ at $t \sim 0.1 \div 1h$); then, in a time scale which is two order of magnitudes larger, the polymers do colocalize ($p \sim 100\%$ at $t \gtrsim 10h$) and the molecular cluster includes both of them. **Panel B** These 2D projections of the system show the configurations at three different times, with the polymer center-of-mass marked by green circles and a colour map representing the density of type-2 molecules (red colour indicates high density regions, blue low density). At $t = 0$ (upper panel), the polymers are far apart and the molecules are quite uniformly distributed in the lattice; at $t = 0.1h$, the polymers are still far from each other, but a unique molecular cluster is formed around one of the two polymers; finally, at $t = 10h$, the polymers are paired and the molecular cluster includes both of them.

# Chapter 4

# A possible scenario for X Inactivation in real cells

In this final chapter, we will use the results of chapter 2 and 3 to try to depict a comprehensive scenario for X Chromosome Inactivation, which includes the current experimental observations in Wild Type cells (i.e., normal cells, without any mutations) as well as in mutants (say, cells with DNA deletions/insertions and cells with multiple copies of X chromosomes; see chapter 1 for details on the mutations) [72, 9] .

## 4.1 "Wild Type" female and male cells

Summarizing, in the previous chapter, we considered a model system where two kinds of molecular factors interact with a couple of identical polymers and naturally establish an architecture where two hubs (composed of type-a and type-b regions in contact with type-c region) are spontaneously formed. Such a spatial conformation is produced by a switch-like thermodynamic mechanisms (a colocalization transition) as soon as molecule concentration/DNA-binding energy rises above a threshold value. Upon activation of mutual interactions of the molecules within each class, the architecture of the two polymers is spontaneously driven into opposite directions ("Symmetry Break-

ing"): the type-a hub persists on one, randomly selected, polymer where the type-b hub is open, whereas the opposite conformations occurs on the other polymer. The mutual interactions between A (and B) molecules could be induced, e.g., by chemical modifications or by additional factors which facilitate their association.

The symmetry breaking mechanism exploits the elements involved earlier in the formation of type-a and type-b hubs, but it originates from a different thermodynamic phase transition: the mutual interactions between molecules induce the self-assembling of a single supermolecular aggregate for each class which randomly, yet in a mutually exclusive way (for the competition for the type-c region), binds to one of its equivalent polymer targets. This process is also switch-like and results in a new sharp, yet stochastic, regulatory mechanism.

We propose that the early stages of XCI could be governed by physical mechanisms of "colocalization" and "symmetry breaking" like those illustrated in the previous chapters: the connection between architectural changes and choice of fate would be naturally rationalized; and their on-off character would explain how "counting&choice" are reliably regulated.

Upon initiation of XCI, the $Jpx$ and $Xist$ genes colocalize with the "buffer" region in the so-called ACH1 hub, while $Tsix$, $Xite$ and the "buffer" form the ACH2 hub (see section 1.4 and [85]). Our model can explain, on simple physical grounds, how those elements can be sharply and simply regulated to recognize each other and colocalize. Later in XCI the opposite fate of X chromosomes is established at random. And we showed how the same physical elements involved in the hub formation process can spontaneously break the X symmetry: the molecular aggregate bound, in our model, to the $Tsix$-$Xite$ locus (type-b loci) is interpreted as a factor related to $Xist$ silencing (i.e., to its Blocking Factor [4, 91, 46]) and designates the future active X; the different aggregate bound to the $Jpx$-$Xist$ locus (type-a loci) of the other X, is linked to an $Xist$ activating factor [82, 36, 84]. The observed pairing of $Tsix/Xite$ region can also follow from this mechanism (see section

3.5).

The two types of molecular aggregates we envisage in our model, are likely to be made up of both specific factors, such as non-coding RNA (like $Tsix$, $Xite$, $Jpx$, etc.; see [46] and next section), and unspecific factors, i.e., proteins (e.g., $Ctcf$ [12, 92], or other putative factors such as $Yy1$ [25], $Oct4$ [24], etc.).

The time required to break the symmetry and so to assemble the "activating" and the "blocking" factor, can be proved to be a power-law function of the chromosome distance [63]. Thus, it may be necessary for the XCI to occur in a proper time, that the $Xic$ regions where the two factors bind, are brought close one to each other, though not in physical contact. This can be achieved thanks to the pairing of another locus, close to the regions where the hubs are: one possible candidate is the $Xpr$ locus, which pair off before XCI starts (even before the $Tsix/Xite$ pairing; see [3] and section 1.3). The pairing of these loci can be the way the X's "sense" one each other and the symmetry breaking mechanism is triggered.

Interestingly, the model can explain, for the first time in a unified framework, all current major experimental observations on XCI, and, in particular, the effects of deletions in the area (see below and [91, 46, 63, 62, 73, 82]).
A two factors regulated XCI in male cells is naturally framed in the same context too. Indeed, in males, whose single X chromosome must remain active, other processes could intervene, yet it is easy to see how the same two factors mechanism can work, i.e., why the only X is usually bound by the B aggregate (and not by A) to repress $Xist$. In fact, the affinities of A and B molecules for the key type-c region are expected, in general, to be different: $E_A \neq E_B$. Hence, if $E_B$ is larger than $E_A$, it is thermodynamically convenient that B molecules bind the X, a difference of a few units in $kT$ being sufficient to skew of orders of magnitudes the binding probability of A and B.

The type-c region in our model emerges as a key control cis-regulator which orchestrates productive contacts by binding one or the other hub. Its

key role in XCI could be tested by experiments targeted at the "buffer" area. Our model also predicts non-linear threshold effects in, e.g., genetic deletions along the regulatory regions which, according to our scenario, master architectural conformations changes and choice of fate. Finally, two main molecular elements are envisaged to control the process and result in the assembling of an activating and a blocking factor for $Xist$.

We considered here a version of the model including just minimal ingredients, as it can better serve the purpose to illustrate the key mechanisms of regulation. Yet, the model can accommodate further elements and more precise molecular details. More generally, the simplicity and robustness of thermodynamic switches like those described here could make them relevant to other nuclear processes, e.g., chromatin architectural organization (see also chapter 2 and [40, 55, 27]) and random monoallelic expression (see below and [30]).

In fact, XCI is only one of the best studied examples of an entire class of cellular processes which involve stochastic regulatory mechanisms. Such processes are known as "Random Monoallelic Expression". In diploid organisms (see note on page 12) it is generally assumed that the paternally and paternally derived copies of each gene (called "alleles") are simultaneously expressed at comparable levels. However there are important exceptions where only one of the two alleles is expressed, and in several cases the choice of the allele to express is made at random by the cell. The number of genes in humans which undergo such "random monoallelic expression", according to recent estimations, is about the 10% of all human genes [30], and include classes of genes which are fundamental, e.g., for the olfactory and the immune system. And in all these cases the underlying physical mechanisms remain unknown. XCI is a somewhat peculiar example in the fact that it entails the silencing of an entire chromosome. Yet, the model that we described, as it is very general and uses a few basic molecular ingredients, can be as well relevant to explain these other examples of cellular "random choices".

## 4.2 Deletion and insertion experiments

As we anticipated, with the model we propose, it is also possible to rationalize the effects on XCI of the deletion/insertion experiments reported in literature. In our perspective, a deletion in a DNA segment, including binding sites for the molecular components of the "Blocking Factor" (BF) or the "Activating Factor" (AF; see previous section), results in the reduction of the chemical affinity of such a sequence for the BF or the AF; analogously, the insertion of a similar segment into an autosome results in the possibility of the autosome to bind the BF or the AF. So, with reference to the data reported in the previous chapter, we can interpret, on a quantitative ground, deletion and insertion experiments.

For instance, let us consider the deletions illustrated in the section 1.4, whose outcomes we summarize here. The heterozygous deletion $\Delta 65kb$ [15] turns out to be lethal in males, as it leads to the inactivation of the only X, while in females it always determines the inactivation of the deleted X. The smaller deletions ($\Delta AS$, $\Delta AJ$, $\Delta AV$, $\Delta 34$ [86], $Tsix^{\Delta CpG}$ [47] and $Xite^{\Delta L}$ [65]), nested in the $\Delta 65kb$ deletion, may cause the X inactivation in male cells only in a certain fraction of cases. The deletions $Tsix^{\Delta CpG}$ and $Xite^{\Delta L}$ result in the inactivation of the deleted X in heterozygous females. A random choice but a "chaotic" counting is found in homozygous deletions in females, with a number of cells showing two active X chromosomes instead of one. The insertions of this regions in non-sex chromosome of male cells cause the inactivation of the only X.

All these results can be explained by our model, since the deleted areas involve the DNA sites where the binding sites for the "Blocking Factor" (which prevents the inactivation of the X it binds to) are mapped (see previous section). The longer deletion $\Delta 65kb$ removes a very large portion of the binding sites, resulting in a very low or a zero affinity for the BF, so it leads to the inactivation of the X that bears the deletion. Yet, the smaller nested deletions are expected to remove a minor fraction of the BF binding sites; so, a "skewed" inactivation is observed in heterozygous deletions, as the BF

will preferentially bind to the Wild Type X, protecting it from inactivation. In homozygous deletions the choice is still random because the BF has the same (lower) affinity for the two X's, but it fails to bind in a fraction of cases, determining the "chaotic" counting. The same holds in males: the only X has no competitors for the BF binding, but there is a probability that BF misses its target and that the only X is inactivated. Such a probability will be higher the more the BF binding sites that are removed, so it is expected to increase with the length of the deletion (as it is experimentally observed). The autosomal insertions of these regions in males allow the BF to bind to the autosome, which competes with the X for it, and may leave the real X chromosome prone to inactivation.

We reported in the section 1.4 the results of some experiments about manipulations of the *Jpx* gene and RNA molecule [84]. Biologists tested the effect of the decrease of *Jpx* RNA molecule concentration and observed that this blocks the XCI in female cells, while it has no effect in males. This can be justified within the symmetry breaking model, by assuming that *Jpx* RNA enters in the formation of the envisaged AF: in this case, the elimination of this molecule may prevent the AF assembling, and consequently, the inactivation of one X.

The heterozygous deletion of *Jpx* gene, $\Delta Jpx$, has no effect in males, whereas it is found to be lethal for female cells at least in 85% of cases, as the inactivation gets blocked. And, the few female cells which survive this mutation, are shown to preferentially inactivate the Wild Type X chromosome (section 1.4 and [84]).

In the previous section, we saw that at least some of the binding sites for the AF are mappable around the *Jpx* gene. Thus, an heterozygous deletion of *Jpx* in a female cell could determine an inability of the AF to bind to the mutated X chromosome and to trigger the inactivation of it. For this reason, a failure of the X inactivation in about the $\sim$ 50% of cases, when the BF binds to the Wild Type X and the mutated X is designated to get inactivated, is expected. The percentage of failed XCI can be even larger if

it is considered that the $Jpx$ RNA molecule enters in the formation of the AF, as the above discussed experiment suggests. Summarizing, the $\Delta Jpx$ deletion, has a double effect that tends to block the inactivation of the X: it hinders the AF assembling because it decreases the concentration of one of its component, and makes the AF binding to the mutated X more difficult (and this explains why the survived cells inactivate more often the Wild Type X). The absence of major effects of the same deletion in male cells is easily rationalized: as in males the BF is the only molecular complex which binds to the only X (see previous section), a mutation like $\Delta Jpx$ which affects the formation/binding of the AF is going to be harmless.

Another very interesting experimental result was obtained by a double heterozygous mutations in females: if the lethal $\Delta Jpx$ mutation is accompanied by a truncation of $Tsix$ gene on the same X, the cells are able to carry out the X inactivation and survive [84].
In the view of our model, $Tsix$ hosts some binding sites for the BF and is a possible component of the BF itself as a RNA molecule (see previous section). Thus, the truncation of $Tsix$ gene is expected to affect the BF assembling/binding approximately as the $Jpx$ deletion does with the AF. Yet, since according to our symmetry breaking model, the AF or the BF assembling on a X chromosome results from the competition between the respective molecular components on their binding sites, the effect of the double mutation may be different from the mere "sum" of the single mutations effects. In fact, let's consider, for instance the AF: while the $\Delta Jpx$ reduces the affinity of the AF for the deleted X, the $Tsix$ truncation tends to favour its binding since it reduces the competition due to the BF binding. So, one effect can counterbalance the other and rescue the lethal phenotype of the $\Delta Jpx$ single mutation.

## 4.3 Multiploid cells

So far, in our models, we aim at describing X inactivation in female cells, i.e., cells with one couple of X chromosomes. However, on the bases of our results, we can speculate on what happens in cells with multiple copies of X. In the scenario we depicted for the X inactivation, the symmetry breaking mechanism and so the "counting&choice" of the X chromosomes is triggered whenever the *Xic* regions are brought in apposition, for instance, by the pairing of *Xpr* locus (see above section 4.1). One speculation is that the outcome of the XCI in multiploid cells is linked to the pairing configuration of the X chromosomes, as only paired groups of X's are able to "sense" each other and to interact via the symmetry breaking mechanism. So, a paired couple of X's would result in the inactivation of just one X, like in normal female cells. Conversely, if an unpaired X is present in the system, this remains active, such as in male cells. In a group of more than two X's, what is expected to happen at equilibrium, is that the AF and the BF assemble around two X's leaving all the other X's "naked" in the open configuration (see fig. 4.1). As this open configuration has been found to correspond to active X chromosomes [85], in a group of, say, $n$ paired chromosomes, we expect to have only one X inactivated. Figures 4.1 and 4.2 summarize these predictions respectively in cells with 4 and 3 X's, by illustrating for each pairing configuration, the predicted spatial architecture at *Xic* and the number of inactive X's.

We showed in the second chapter that the pairing of distant DNA loci can be produced by a "passive shuttling" mechanism. Under the assumption that the pairing of X chromosomes is produced by such a mechanism, we can use the analyses we carried out in the section 2.5.1.4 to predict the probability to find each possible pairing configuration in cells with 3 and 4 copies of X chromosomes. Then, the previous discussion can be used to map the probability distribution of pairing into, e.g., the probability distribution of the number of inactive X to check our predictions against the experimental data coming from XXXX and XXXY cells (see section 1.6). Some precautions must be taken in doing such a comparison: while our data are referred to an

ideal population of cells which are perfectly syncronized, the experimental samples do contain cells at different stages of differentiation and of XCI process. The probability distribution measured in real cells is also affected by the death of cells with a wrong number of inactive X and by cell divisions, occurring at a rate which depends on the number of inactive X [57]. However, at least a qualitative check against the experiments can be made, by looking at the shapes of the distribution.

Let us start with cells having four X chromosomes. We found different scenarios depending on the details of the interaction between the molecular mediators and the DNA loci (see 2.5.1.4). In a first model, where the same DNA sequence can form multiple bonds with the molecular mediators, two thermodynamics phases are observed, with the four DNA segments which either diffuse independently at equilibrium or, above thresholds in the molecule concentration/affinity, are all paired together. According to the previous discussion (see fig. 4.1), these two pairing configurations correspond to a situation where all the 4 X are active (if the X are all uncoupled) or only one is inactive (when all the 4 X are coupled together). Both the predictions do not agree with the experimental results about X inactivation in XXXX cells (fig. 4.3 green bars in the upper panels).

In a second scenario, where each DNA sequence can bind only to a single molecule at a time, a more complex situation is found: by changing the molecule concentration/affinity three different regimes corresponding to three different probability distribution of the pairing configurations are possible (see fig. 2.11 panel A and B). In the upper panels of fig. 4.1, the predicted distribution $\mathcal{P}$ of the number of inactive X in each of the three regimes are plotted (blue bars) together with the experimental results in XXXX cells (green bars). An excellent agreement is found in the phase where most of the cells present two independent couples of paired X (upper right panel).

It is possible to repeat the same analysis for cells with 3X. In this case, we saw that the pairing configurations do not depend on the details of the DNA-molecule interaction and three possible pairing distributions are found (see

section 2.5.1.4 and fig. 2.11 panel C and D). The low panels in fig. 4.3 illustrate with blu bars the predicted $\mathcal{P}$ and with purple bars the experimental results from XXXY cells. The predictions are shown to reproduce very well the experimental data both in the configurations where only a couple of X is formed (middle low panel) or when all the 3 X are paired (right low panel).

Summarizing, by making some assumptions about the counting&choice step of XCI in the multiploid cells (namely, that X chromosome pairing is a required step for counting&choice and that in a group of paired X's only one is designated to remain active, according to the symmetry breaking model), we obtained from our distributions of paired X, the probability distributions of the number of inactive X's.

By comparing our "in silico" data with the experimental data from XXXX and XXXY cells, we predicted that in XXXX cells two independent X couples are built, while in XXXY cells at least one couple of X must pair off.

These considerations provide, under a number of simplifying hypotheses, quantitative scenarios on how the proposed symmetry breaking model for the "counting&choice" work in multiploid cells. They may also help to find out the relation between $Xic$ pairing and "counting&choice", one of the key issues in X Inactivation.

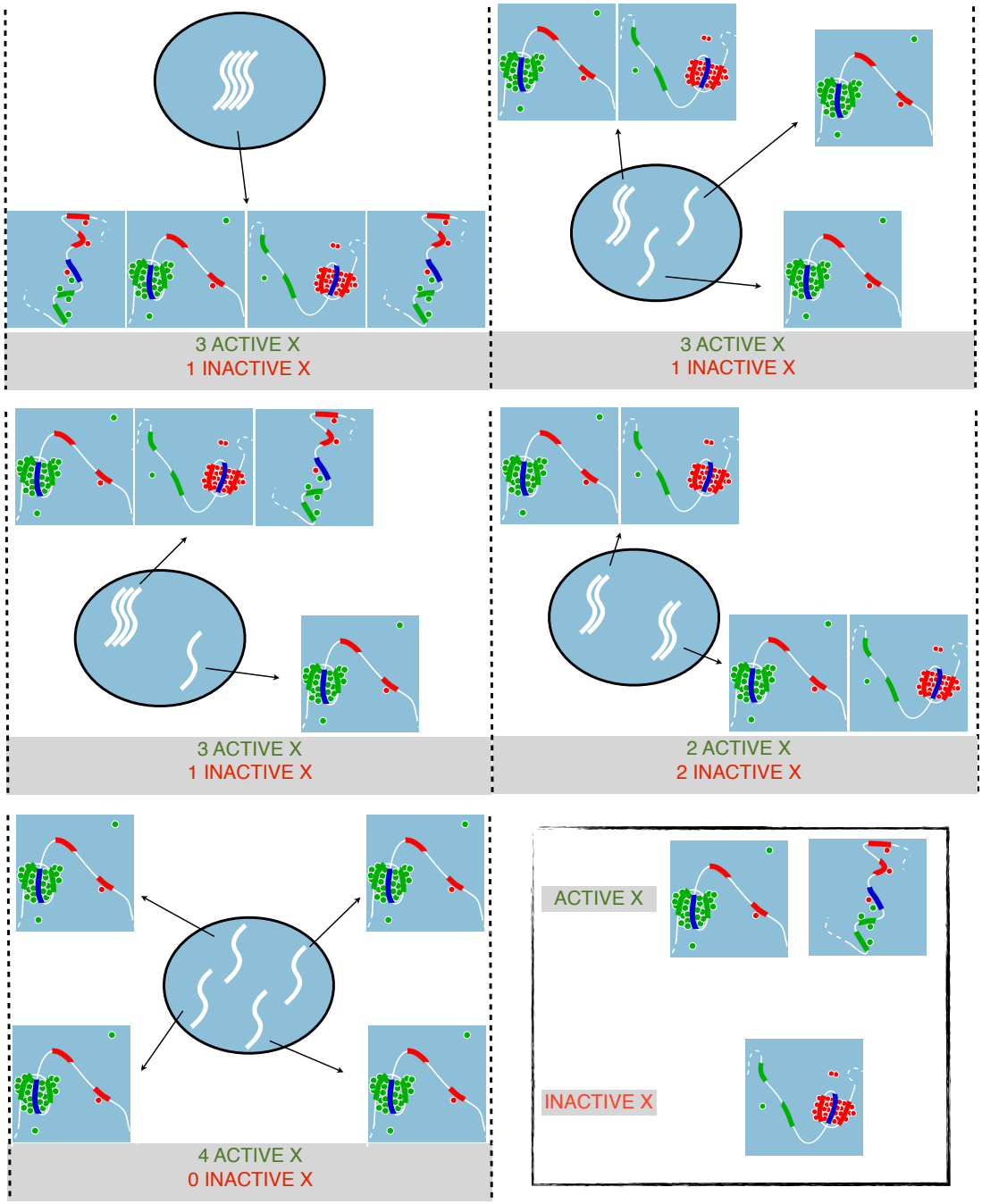


Figure 4.1: By making some assumptions (see text), we estimated the number of active and inactive X's in XXXX cells as function of the pairing configuration of the X chromosomes.

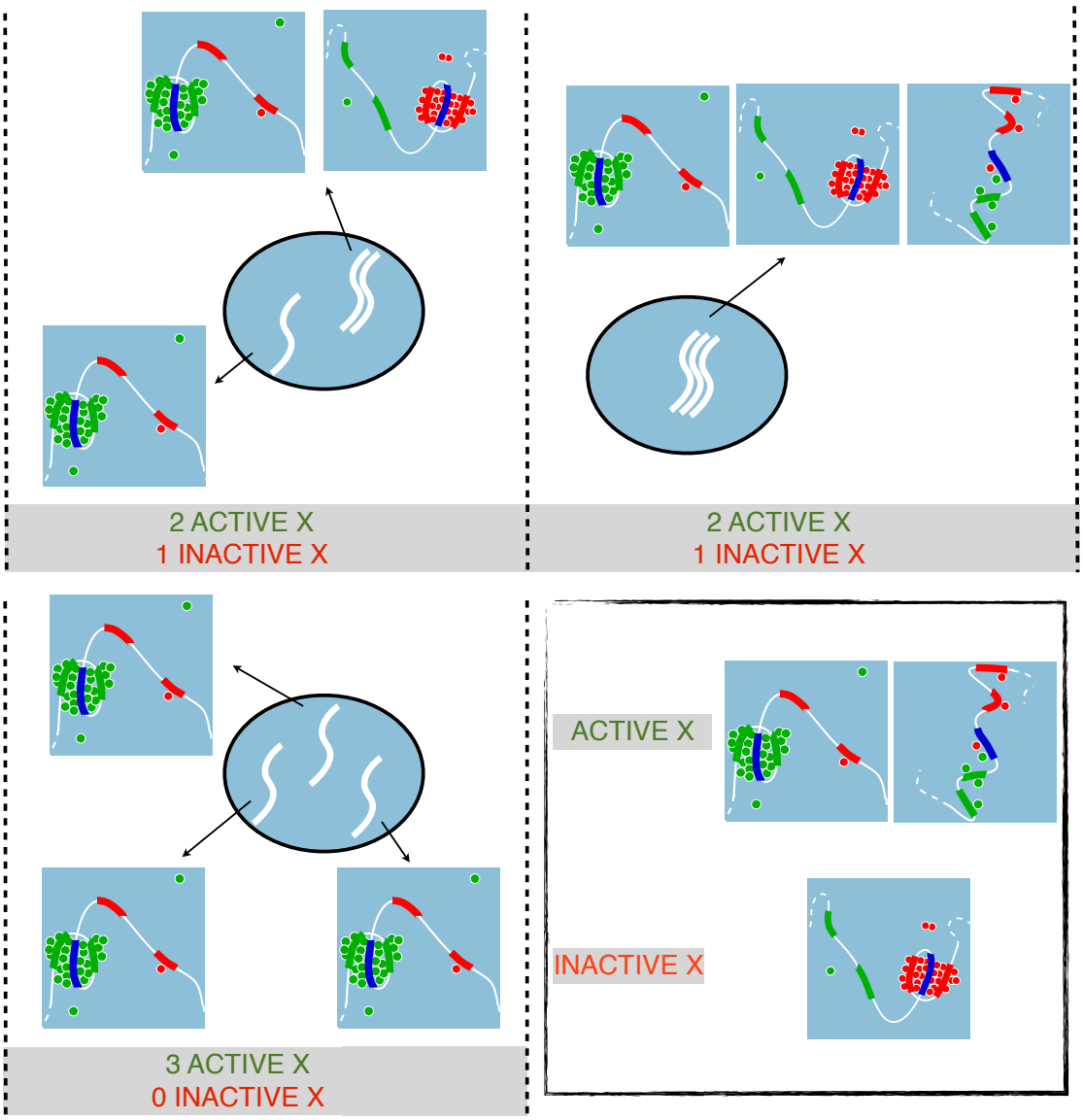


Figure 4.2: Number of active and inactive X's in cells with 3 X chromosomes as function of the pairing configuration.

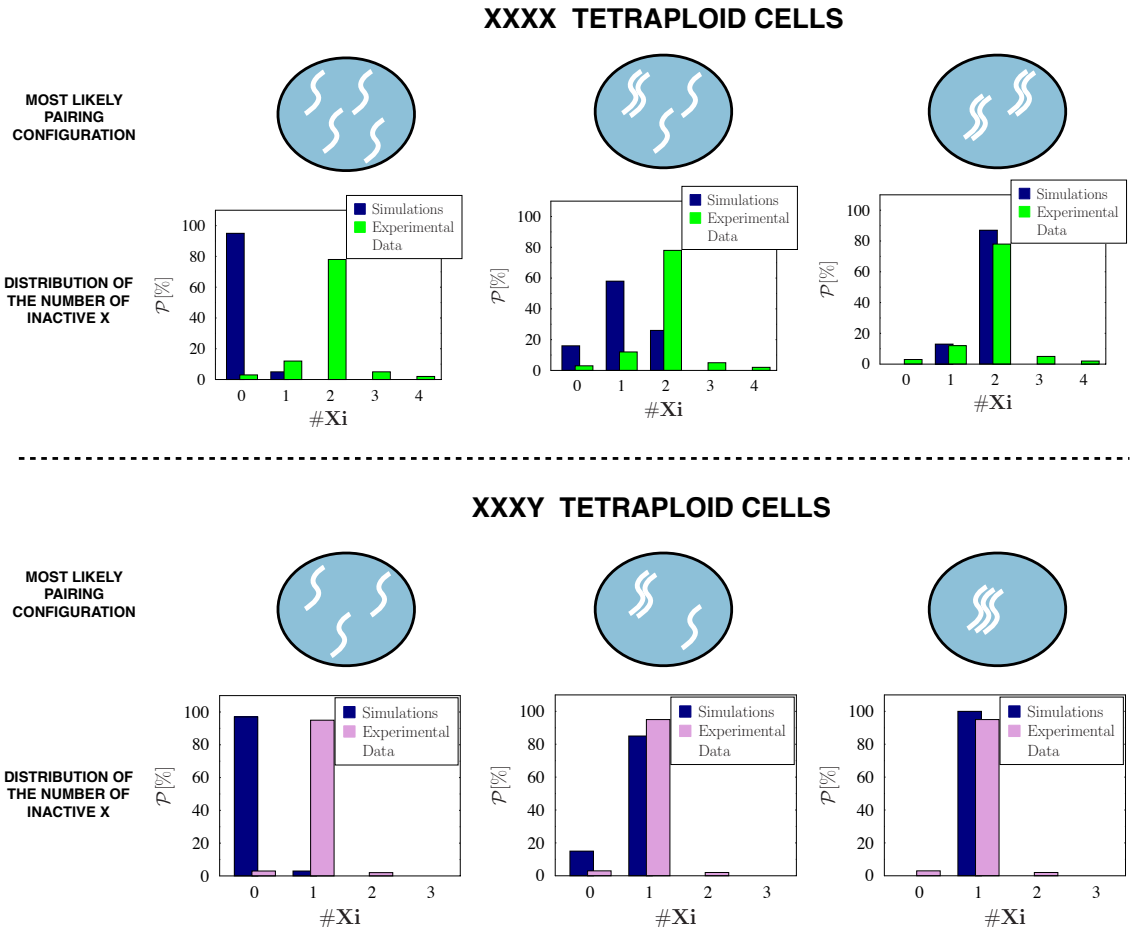


Figure 4.3: In section 2.5.1.4 we predicted the distribution probabilities of the pairing configurations of 3 and 4 DNA loci as produced by the "passive shuttling". We pointed out that different distribution probabilities were found, each characterized by a most likely pairing configuration, that correspond to different thermodynamic phases in the system. We mapped these probability distributions into a probability distribution of the number of inactive X, $\mathcal{P}$, (see text and fig.s 4.1, 4.2) and compare the predictions (blue bars) with the experimental data from XXXX and XXXY cells (green and purple bars)

# Bibliography


[1] R. Agrelo and A. Wutz. X inactivation and disease. *Semin Cell Dev Biol*, 21(2):194–200, Apr 2010.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, 5 edition, November 2007.

[3] S. Augui, G. J. Filion, S. Huart, E. Nora, M. Guggiari, M. Maresca, A. F. Stewart, and E. Heard. Sensing X chromosome pairs before X Inactivation via a novel X-pairing region of the Xic. *Science*, 318(5856):1632–1636, Dec 2007.

[4] P. Avner and E. Heard. X-Chromosome Inactivation: counting, choice and initiation. *Nat Rev Genet*, 2(1):59–67, Jan 2001.

[5] C. P. Bacher, M. Guggiari, B. Brors, S. Augui, P. Clerc, P. Avner, R. Eils, and E. Heard. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol*, 8(3):293–299, Mar 2006.

[6] J. Berg. Dynamics of gene expression and the regulatory inference problem. *Europhys. Lett.*, 82:28010, 2008.

[7] O. G. Berg and P. H. von Hippel. Selection of DNA binding sites by regulatory proteins. *Trends Biochem Sci*, 13(6):207–11, Jun 1988.

[8] O. G. Berg, R. B. Winter, and P. H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, 20(24):6929–6948, Nov 1981.

[9] V. Bianco, M. Nicodemi, and A. Scialdone. *in preparation*, 2010.

[10] K. Binder. Applications of Monte Carlo methods to statistical physics. *Reports on Progress In Physics*, 60(5):487–559, MAY 1997.

[11] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987.

[12] W. Chao, K. D. Huynh, R. J. Spencer, L. S. Davidow, and J. T. Lee. CTCF, a candidate trans-acting factor for X-inactivation choice. *Science*, 295(5553):345–7, Jan 2002.

[13] C.-H. Chuang, A. E. Carpenter, B. Fuchsova, T. Johnson, P. de Lanerolle, and A. S. Belmont. Long-range directional movement of an interphase chromosome site. *Curr Biol*, 16(8):825–831, Apr 2006.

[14] J. R. Chubb, S. Boyle, P. Perry, and W. A. Bickmore. Chromatin motion is constrained by association with nuclear compartments in human cells. *Current Biology*, 12(6):439–45, Mar 2002.

[15] P. Clerc and P. Avner. Role of the region 3' to Xist exon 6 in the counting process of X-chromosome inactivation. *Nat Genet*, 19(3):249–53, Jul 1998.

[16] D. E. Cohen, L. S. Davidow, J. A. Erwin, N. Xu, D. Warshawsky, and J. T. Lee. The DXPas34 repeat regulates random and imprinted X inactivation. *Dev Cell*, 12(1):57–71, Jan 2007.

[17] P. G. de Gennes. *Scaling concepts in polymer physics*. Cornell University Press, Ithaca, N.Y., 1979.

[18] W. de Laat. Long-range DNA contacts: romance in the nucleus? *Curr. Op. Cell Bio.*, 19:317–320, 2007.

[19] W. de Laat and F. Grosveld. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*, 11(5):447–459, 2003.

[20] K. DeBell and T. Lookman. Surface phase-transitions in polymer systems. *Reviews of Modern Physics*, 65(1):87–113, 1993.

[21] J. Dekker. Gene regulation in the third dimension. *Science*, 319(5871):1793–1794, Mar 2008.

[22] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–11, Feb 2002.

[23] M. Doi and S. Edwards. *The Theory of Polymer Dynamics*. Clarendon Press, Oxford, USA, 1986.

[24] M. E. Donohoe, S. S. Silva, S. F. Pinter, N. Xu, and J. T. Lee. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature*, 460(7251):128–32, Jul 2009.

[25] M. E. Donohoe, L.-F. Zhang, N. Xu, Y. Shi, and J. T. Lee. Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell*, 25(1):43–56, Jan 2007.

[26] M. Dundr, J. K. Ospina, M.-H. Sung, S. John, M. Upender, T. Ried, G. L. Hager, and A. G. Matera. Actin-dependent intranuclear repositioning of an active gene locus in vivo. *J Cell Biol*, 179(6):1095–1103, Dec 2007.

[27] P. Fraser and W. Bickmore. Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447(7143):413–417, May 2007.

[28] S. Galande, P. K. Purbey, D. Notani, and P. P. Kumar. The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr Opin Genet Dev*, 17(5):408–414, Oct 2007.

[29] U. Gerland, J. D. Moroz, and T. Hwa. Physical constraints and functional characteristics of transcription factor-dna interaction. *Proc Natl Acad Sci U S A*, 99(19):12015–12020, Sep 2002.

[30] A. Gimelbrant, J. N. Hutchinson, B. R. Thompson, and A. Chess. Widespread monoallelic expression on human autosomes. *Science*, 318(5853):1136–40, Nov 2007.

[31] L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453:948–951, Jun 2008.

[32] E. Heard, F. Mongelard, D. Arnaud, and P. Avner. Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. *Mol. Cell. Biol.*, 19:3156–3166, Apr 1999.

[33] S. Imai, S. Nishibayashi, K. Takao, M. Tomifuji, T. Fujino, M. Hasegawa, and T. Takano. Dissociation of Oct-1 from the nuclear peripheral structure induces the cellular aging-associated collagenase gene expression. *Mol. Biol. Cell*, 8:2407–2419, Dec 1997.

[34] H. Jing, C. R. Vakoc, L. Ying, S. Mandat, H. Wang, X. Zheng, and G. A. Blobel. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol. Cell*, 29:232–242, Feb 2008.

[35] J. F. Joanny, L. Leibler, and P. G. de Gennes. Effects of polymer solutions on colloid stability. *J. Polym. Sci., Polym. Phys.*, 17:1073–1084, 1979.

[36] I. Jonkers, T. S. Barakat, E. M. Achame, K. Monkhorst, A. Kenter, E. Rentmeester, F. Grosveld, J. A. Grootegoed, and J. Gribnau. RNF12

is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell*, 139(5):999–1011, Nov 2009.

[37] A. B. Kolomeisky and M. E. Fisher. Molecular motors: a theorist's perspective. *Annu Rev Phys Chem*, 58:675–95, 2007.

[38] P. P. Kumar, O. Bischof, P. K. Purbey, D. Notani, H. Urlaub, A. Dejean, and S. Galande. Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat. Cell Biol.*, 9:45–56, Jan 2007.

[39] R. I. Kumaran, R. Thakar, and D. L. Spector. Chromatin dynamics and gene positioning. *Cell*, 132(6):929–934, Mar 2008.

[40] 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. 8:104–115, 2007.

[41] L. D. Landau and E. M. Lifshitz. *Statistical Physics: Volume 5*. Butterworth-Heinemann, 1980.

[42] C. Lanzuolo, V. Roure, J. Dekker, F. Bantignies, and V. Orlando. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat Cell Biol*, 9(10):1167–1174, Oct 2007.

[43] M. Lassig. From biophysics to evolutionary genetics: statistical aspects of gene regulation. *BMC Bioinformatics*, 8 Suppl 6:S7, 2007.

[44] J. T. Lee. Homozygous Tsix mutant mice reveal a sex-ratio distortion and revert to random X-inactivation. *Nat Genet*, 32(1):195–200, Sep 2002.

[45] J. T. Lee. Regulation of X-chromosome counting by Tsix and Xite sequences. *Science*, 309(5735):768–71, Jul 2005.

[46] J. T. Lee. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev*, 23(16):1831–42, Aug 2009.

[47] J. T. Lee and N. Lu. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell*, 99(1):47–57, Oct 1999.

[48] J. T. Lee, W. M. Strauss, J. A. Dausman, and R. Jaenisch. A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell*, 86(1):83–94, Jul 1996.

[49] S. Lin and A. D. Riggs. The general affinity of lac repressor for E. coli DNA: implications for gene regulation in procaryotes and eukaryotes. *Cell*, 4(2):107–11, Feb 1975.

[50] D. M. Lonard and B. W. O'Malley. Gene transcription: two worlds merged. *Nature*, 452(7190):946–947, Apr 2008.

[51] S. J. Maerkl and S. R. Quake. A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, 315(5809):233–7, Jan 2007.

[52] C. Massie and I. Mills. Chipping away at gene regulation. *EMBO Rep. Review*, 9:337–43, 2008.

[53] K. Meaburn and T. Misteli. Cell biology: chromosome territories. *Nature*, 445:379–781, 2007.

[54] T. Misteli. Protein dynamics: implications for nuclear architecture and gene expression. *Science*, 291(5505):843–847, Feb 2001.

[55] T. Misteli. Beyond the sequence: cellular organization of genome function. *Cell*, 128(4):787–800, Feb 2007.

[56] T. Misteli. Cell biology: Nuclear order out of chaos. *Nature*, 456(7220):333–334, Nov 2008.

[57] K. Monkhorst, I. Jonkers, E. Rentmeester, F. Grosveld, and J. Gribnau. X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. *Cell*, 132(3):410–421, Feb 2008.

[58] C. Morey, P. Navarro, E. Debrand, P. Avner, C. Rougeulle, and P. Clerc. The region 3' to Xist mediates X chromosome counting and H3 Lys-4 dimethylation within the Xist gene. *EMBO J*, 23(3):594–604, Feb 2004.

[59] A. V. Morozov, J. J. Havranek, D. Baker, and E. D. Siggia. Protein-DNA binding specificity predictions with structural models. *Nucleic Acids Res*, 33(18):5781–98, 2005.

[60] M. Nicodemi, B. Panning, and A. Prisco. Colocalization transition of homologous chromosomes at meiosis. *Phys. Rev. E*, 77:061913, 2008.

[61] M. Nicodemi, B. Panning, and A. Prisco. A thermodynamic switch for chromosome colocalization. *Genetics*, 179(1):717–21, May 2008.

[62] M. Nicodemi and A. Prisco. Self-assembly and DNA binding of the blocking factor in X chromosome inactivation. *PLoS Comput Biol*, 3(11):e210, Nov 2007.

[63] M. Nicodemi and A. Prisco. Symmetry-breaking model for X-chromosome inactivation. *Phys Rev Lett*, 98(10):108104, Mar 2007.

[64] M. Nicodemi and A. Prisco. Thermodynamic pathways to genome spatial organization in the cell nucleus. *Biophys J*, 96(6):2168–2177, Mar 2009.

[65] Y. Ogawa and J. T. Lee. Xite, X-Inactivation intergenic transcription elements that regulate the probability of choice. *Mol Cell*, 11(3):731–43, Mar 2003.

[66] R.-J. Palstra, B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, and W. de Laat. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet*, 35(2):190–194, Oct 2003.

[67] B. Payer and J. T. Lee. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet*, 42:733–772, 2008.

[68] R. D. Phair and T. Misteli. High mobility of proteins in the mammalian cell nucleus. *Nature*, 404(6778):604–9, Apr 2000.

[69] E. M. Pugacheva, V. K. Tiwari, Z. Abdullaev, A. A. Vostrov, P. T. Flanagan, W. W. Quitschke, D. I. Loukinov, R. Ohlsson, and V. V. Lobanenkov. Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. *Hum Mol Genet*, 14(7):953–65, Apr 2005.

[70] W. W. Quitschke, M. J. Taheny, L. J. Fochtmann, and A. A. Vostrov. Differential effect of zinc finger deletions on the binding of CTCF to the promoter of the amyloid precursor protein gene. *Nucleic Acids Res*, 28(17):3370–3378, Sep 2000.

[71] M. Renda, I. Baglivo, B. Burgess-Beusse, S. Esposito, R. Fattorusso, G. Felsenfeld, and P. V. Pedone. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J Biol Chem*, 282(46):33336–45, Nov 2007.

[72] A. Scialdone, I. Cataudella, A. Prisco, and M. Nicodemi. *in preparation*, 2010.

[73] A. Scialdone and M. Nicodemi. Mechanics and dynamics of X-chromosome pairing at X Inactivation. *PLoS Comput Biol*, 4(12):e1000244, Dec 2008.

[74] A. Scialdone and M. Nicodemi. DNA Loci Cross-Talk through Thermodynamics. *J Biomed Biotechnol*, 2009:516723, 2009.

[75] A. Scialdone and M. Nicodemi. Diffusion-based DNA target colocalization by thermodynamic mechanisms. *Development*, 137(22):3877–85, Nov 2010.

[76] A. Scialdone and M. Nicodemi. Passive DNA shuttling. *Europhysics Letters*, 92(2):20002, 2010.

[77] A. Scialdone and M. Nicodemi. Statistical Mechanics models for X-Chromosome Inactivation. *Advances in Complex Systems*, 13(3):367–376, JUN 2010.

[78] A. Semenov, J. Avalos, A. Johner, and J. Joanny. Adsorption of polymer solutions onto a flat surface. *Macromolecules*, 29(6):2179–2196, 1996.

[79] M. Slutsky and L. A. Mirny. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys J*, 87(6):4021–4035, Dec 2004.

[80] J. A. Spudich. The myosin swinging cross-bridge model. *Nat Rev Mol Cell Biol*, 2(5):387–92, May 2001.

[81] H. E. Stanley. *Introduction to phase transitions and critical phenomena*. Clarendon Press, Oxford, USA, 1971.

[82] J. Starmer and T. Magnuson. A new model for random X chromosome inactivation. *Development*, 136(1):1–10, Jan 2009.

[83] T. Takizawa, K. J. Meaburn, and T. Misteli. The meaning of gene positioning. *Cell*, 135(1):9–13, Oct 2008.

[84] D. Tian, S. Sun, and J. T. Lee. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell*, 143(3):390–403, Oct 2010.

[85] C.-L. Tsai, R. K. Rowntree, D. E. Cohen, and J. T. Lee. Higher order chromatin structure at the X-Inactivation center via looping DNA. *Dev Biol*, 319(2):416–25, Jul 2008.

[86] S. Vigneau, S. Augui, P. Navarro, P. Avner, and P. Clerc. An essential role for the DXPas34 tandem repeat and Tsix transcription in the

counting process of X chromosome inactivation. *Proc Natl Acad Sci U S A*, 103(19):7390–5, May 2006.

[87] P. H. von Hippel and O. G. Berg. On the specificity of DNA-protein interactions. *Proc Natl Acad Sci U S A*, 83(6):1608–12, Mar 1986.

[88] J. D. Watson, T. A. Baker, S. P. Bell, A. Gann, M. Levine, and R. Losick. *Molecular Biology of the Gene*. Benjamin Cummings, 5th edition, December 2003.

[89] R. B. Winter, O. G. Berg, and P. H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor–operator interaction: kinetic measurements and conclusions. *Biochemistry*, 20(24):6961–6977, Nov 1981.

[90] R. B. Winter and P. H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor–operator interaction: equilibrium measurements. *Biochemistry*, 20(24):6948–6960, Nov 1981.

[91] A. Wutz and J. Gribnau. X inactivation Xplained. *Curr Opin Genet Dev*, 17(5):387–393, Oct 2007.

[92] N. Xu, M. E. Donohoe, S. S. Silva, and J. T. Lee. Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet*, 39(11):1390–1396, Nov 2007.

[93] N. Xu, C.-L. Tsai, and J. T. Lee. Transient homologous chromosome pairing marks the onset of X inactivation. *Science*, 311(5764):1149–1152, Feb 2006.

[94] L. F. Zhang, K. D. Huynh, and J. T. Lee. Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell*, 129:693–706, May 2007.

[95] Y. Zhao, D. Granas, and G. D. Stormo. Inferring binding energies from selected binding sites. *PLoS Comput Biol*, 5(12):e1000590, Dec 2009.