Università degli Studi di Napoli "Federico II"

Doctoral Thesis

Statistical Mechanics Models of Chromatin Architecture

Author:
Mariano Barbieri

Supervisor:
Prof. Mario Nicodemi

April 2013

"Il buon senso era lì, ma se ne stava nascosto per paura del senso comune."

"Good sense was there, but was concealing because scared by common sense."

Alessandro Manzoni

Acknowledgements

Ringrazio tutte le persone che ho conosciuto e tutte le esperienze che ho fatto, tutto ha concorso a rendermi la persona che ha lavorato a questa tesi in questi anni.

Ringrazio la mia famiglia e le sorelle Incoronato, che non smettono mai di trasmettermi il loro carico di umanità e di saggezza profonda della vita, anche se non se ne accorgono. Anzi quando cercano di farlo appositamente non ci riescono.

Ringrazio l'ambiente universitario napoletano, perché, diversamente da quello che si può pensare, è il posto più piacevole che conosca dove lavorare.

Ringrazio il mio supervisor Mario "Mimmo" Nicodemi, che è riuscito a tirarmi fuori qualcosa anche nei periodi non troppo positivi, spero un giorno di esser bravo come lui.

Ringrazio l'aula dottorandi e gli abitanti lì succedutisi in questi anni, perché se ora riesco a fare qualcosa di buono anche in un ambiente sfavorevole è anche grazie all'allenamento a cui loro mi hanno sottoposto.

Alcuni di questi personaggi val la pena di menzionare: Lorenzo detto "'er Puccio", Alan alias "Superman" e le sconclusionate uscite serali in loro compagnia; Fatema "The Lady of the Ladies", Aboul-"Double Bubble Hubble"-Fazl, Sergei "Desmund" e le conclusionatissime serate passate nei loro appartamenti.

Ringrazio Piero Landolfi, Deborah Pallotti e Antonella Bianchi, perché sono state le persone che più hanno dato significato al mio percorso di vita.

Contents

Introduction

Understanding the interplay between genome architecture and gene regulation is one of the most challenging problems in biology. During mitosis, chromosomes are found in a condensed state, but decondense during interphase, when highly coordinated cellular processes such as transcription, DNA repair, and replication take place, creating cell-type-specific chromatin folding [1–3].

Chromosome organization occurs at different scales of genomic length to yield variable degrees of compaction [4]. Linear nucleosome arrays fold into higher-order structures, first through local chromatin interactions, such as between promoters and enhancers, and then eventually giving rise to discrete chromosome territories [1].
Spatial genome organization is guided by intra- and inter-chromosomal interactions mediated by nuclear components that include transcription factors, transcription and replication factories, Polycomb bodies, and contacts with the lamina [5–8]. However, how binding of diffusible factors to specific genomic regions drives chromatin folding remains poorly understood.

Imaging of single loci by FISH and genome-wide mapping of chromatin interactions by chromosome conformation capture (3C) approaches (see Appendix A) revealed a variety of chromatin architectures across genomic regions and cell types, and upon environmental cues [9–13] (Fig. 1.6A). In FISH experiments, chromatin folding is often measured by the mean-square spatial distance, $R^2(s)$, between two genomic regions as a function of their linear genomic distance, s (Fig. 1.6B), which usually exhibits scaling properties $R^2(s) \sim s^{2\nu}$. Although the behavior of $R^2(s)$ appears to depend on the genomic regions and cell types assessed (Fig. 1.6A), in general, at large genomic distances, $R^2(s)$ reaches a plateau (i.e., $\nu = 0$) that reflects the folding of chromosomes into territories [14].

A global analysis of genome-wide 3C (Hi-C) (Appendix A) ligation products in human cells averaged across all chromosomes has been used to estimate the "contact probability" $P_c(s)$ [12]. This measures how frequently two loci contact each other as a function of s (Fig. 1.6B). Measurements of $P_c(s)$ have identified a power-law behavior, $P_c(s) \sim 1/s^{\alpha}$, with an average exponent $\alpha \sim 1.08$, at genomic distances $0.5 - 7$ Mb.

The observation of α of approximately 1.08 has led to the suggestion that chromatin behavior could be explained by a single folding structure, previously described in polymer physics by the "fractal-globule" (FG) model [15]. Recent applications of Hi-C have found a lower exponent α in Drosophila [13], and revealed differing contact frequencies in human cells according to chromatin expression status [16]. The observation that the exponent α may not be universal (see also Figs. 1.7,1.9) or conserved, and the failure of the FG model to describe the plateauing of $R^2(s)$, prompted us to reconsider the fundamental underlying principles of chromatin folding.

In [17], I explain in Chapter 1, we explored how chromatin architectural patterns can arise and be regulated by using an alternative simple polymer physics model, first proposed in 2008 [18, 19]. In our model, the "strings and binders switch" (SBS) model, nonrandom chromatin conformations are established through attachment of diffusible factors (binders) to binding sites. Binder-mediated interactions give rise to a variety of stable chromatin architectures that can coexist in the nucleus. Chromatin folding changes in response to changes in binding site distribution, binder concentration, or binding affinity, in a switch-like fashion across specific threshold values via thermodynamics mechanisms. Importantly, we show that the SBS model describes in a single framework all current experimental data on chromosome architecture from FISH, Hi-C and 3C approaches (see Appendix A about these experimental techniques).

We already have demonstrated how powerful the ingredients behind SBS model are in a specific cell process and region of genome: the X-Chromosome Inactivation (XCI) phenomenon. In our cells most genes are expressed from both alleles, the most notable exception to this rule being X-linked genes in female mammals. During embryo development, in a step necessary for survival, one randomly selected X chromosome is transcriptionally silenced in each female cell to ensure that the levels of X-derived products are equalized in XX females and XY males. XCI has important scientific and medical implications which have recently focused substantial attention on the mysterious nature of the signals that direct the two chromosomes to opposite fates.[20–22]

XCI is regulated by a 1Mb long region on the X, the Xic (X Inactivation Center), including the Xist gene which is responsible for silencing the presumptive inactive X.[20–22] Xist encodes a non-coding RNA which coats and, thus, triggers silencing of its own X. In female cells, before random XCI initiates, Xist is expressed at low levels from both Xs. Then its expression is upregulated on the future inactive X and switched off on the active X. Silencing of other genes on the Xist-expressing X chromosome follows rapidly.

In [23] (Chapter 2) we investigate the mechanism whereby cells count their Xs and randomly choose the one to inactivate ("counting&choice"). That's one of the most mysterious aspect of XCI:[20–22] how does a cell know that it has two Xs? And how is

the mutually exclusive fate of the two Xist orchestrated?
Biological data suggest that "controlling factors" for counting and choice derive from X and autosomes (i.e., non sex chromosomes) and interact with cis-acting regulatory sequences on the X chromosomes. "Counting&choice" could, thus, be explained via the Blocking Factor (BF) hypothesis:[20–22, 24, 25] the symmetry between the two Xs is broken by a BF which is present in a single copy within the cell nucleus, and binds to the Xic of just one chromosome preventing its inactivation; the second unprotected X is self-inactivated by action of its own Xist gene.

The BF must be a unique entity to perform its function, though, several considerations (e.g., degradation, over-production problems) exclude the possibility that it is a single protein or RNA molecule. On the other hand, if a diffusible controlling factor is produced in several copies that can statistically reach the target, asymmetric binding to two equivalent chromosomes must be explained.

In [26] (Chapter 3) we investigate by computer simulations a schematic model consisting of two identical polymers which interact with a concentration of diffusing molecules (see Fig. 3.1.B). In the light of current Xic 3C data [27], (see Appendix A about 3C technique) the model poses that along each polymer three types of regions exist (type-α, β and γ) and predicts the existence of two types of regulatory molecules (type-A and B).

We show that the system thermodynamic stable states fall in distinct classes corresponding to different conformations. The polymers spontaneously select one of them according to molecule concentration/binding energy. Conformational changes are again driven by thermodynamic phase transitions which act switch-like, regulated by given concentration/binding energy thresholds, as in [17]. The two polymers are exposed to the same environment, yet they can undergo alternative architectural modifications: we show that a symmetry breaking mechanisms is activated if the homotypic interaction between regulatory molecules rises above a threshold.

Comparison to experimental observations [20–22, 27–30] suggests that the regions envisaged by the model can be approximately mapped along the Xic sequence as illustrated in Fig. 3.1.B, while type-A and B complexes could be related to an activating and a blocking regulator of Xist.

Chapter 1

Complexity of chromatin folding is captured by the Strings and Binders Switch model

1.1 Results

1.1.1 The SBS Model: General Description

In the SBS model, a chromatin fiber is represented as a self-avoiding polymer bead chain (Fig. 1.1A), and binding molecules are represented by Brownian particles with concentration c_m. A fraction, f, of polymer sites can be bound by diffusing molecules with chemical affinity E_X . To investigate the system's folding behavior, we evaluated the dynamics and equilibrium properties of the polymer using extensive Monte Carlo (MC) simulations in the known range of the biochemical values of c_m and E_X (see Appendix B for details on simulations). In this instance, we chose a binding multiplicity of diffusing binders equal to six, as estimated for chromatin organizers such as CCCTC-binding factor (CTCF) or transcription factories [31, 32]. This is a representative experimental condition because different binding multiplicities (≥ 2) promote similar patterns of folding [19]. We first demonstrate how c_m affects the equilibrium compaction state of the polymer (Fig. 1.7). The extent of polymer folding is captured by measuring the squared radius of gyration, Rg^2 , which is the average squared distance of each bead to the center of mass of the polymer chain (Fig. 1.6B). R_g^2 attains a minimum when loops enclose the polymer into a compact state, and a maximum when the polymer is loose and randomly folded. The SBS model predicts that $R_g^2(c_m)$ has a sigmoid shape with two distinct regimes and a transition region, as a function of c_m (Fig. 1.7). When c_m

is below a specific threshold value, C_{tr}, R_g^2 has the same value found in the standard random self-avoiding walk (SAW) model [33], where the polymer is open and randomly folded. At the transition point, the conformations of the polymer are fractal [33]. Above threshold, Rg^2 harply decreases towards a value corresponding to a compact, collapsed structure. The threshold, C_{tr} , identified by the curve inflection point, corresponds to the polymer Θ transition [33]. We find, for example, that for $E_X = 2k_BT$, C_{tr} is approximately $10n\text{mol}/L$, a typical nuclear protein concentration [34]. The SBS model therefore explains that a polymer can undergo a switch-like conformational change to form or release loops by changing the concentration (and/or affinity, E_X) of binding molecules. Loops are stable only above C_{tr} , where the system undergoes a thermodynamic phase transition. By changing concentration or affinity across threshold, a thermodynamic switch is controlled to change reliably the polymer architecture (see ref. [19] for more details). The equilibrium folding state of the polymer across a wide range of E_X and c_m values is seen in the system phase diagram (Fig. 1.1B), which shows that the threshold concentration, $C_{tr}(E_X)$, required for switching from open into compact states increases as E_X decreases.

1.1.2 The Conformational Self-Organization Mechanisms of the SBS Model and Its Emerging Stable States

To assess the power of the SBS model in explaining the range of chromatin behaviors observed by FISH and Hi-C, we measured $R^2(s)$, the equilibrium value of the mean-square spatial distance, and the contact probability $P_c(s)$ between loci separated by a distance s along the polymer (Fig. 1.1 C-F). In the SBS model, the shape of the two functions $R^2(s)$ and $P_c(s)$ is sensitive to the concentration of binding molecules, c_m. $R^2(s)$ is characterized by a power-law behavior, $R^2(s) \sim s^{2\nu}$, which defines the scaling exponent ν (at large s; Fig. 1.1 C and D). Importantly, we find that ν is a nonlinear sigmoid function of the concentration of binding molecules, c_m (Fig. 1.1D), which corresponds to a switch-like behavior in the architecture of the polymer. Three regimes exist corresponding to c_m : below, at, and above threshold. Analogous results are found when the binding affinity, E_X, or number of binding sites is varied [19]. When $c_m < C_{tr}$, few chromatin contacts are present and the polymer is open; $R^2(s)$ increases as a function of s with an exponent $\nu \sim 0.58$, as expected for a randomly folded free polymer (the SAW random coil) [33]. If c_m increases, more loops can be formed, but ν remains at approximately 0.58 until C_{tr} is approached. When c_m is around C_{tr}, the polymer architecture changes abruptly from open to a (stable) fractal-like conformation with ν of approximately 0.5. This exponent corresponds to the expected Θ-point exponent of the polymer coil-globule transition [33] and has been observed by FISH [35] (Fig. 1.6A).

When $c_m < C_{tr}$, chromatin contacts are abundant and the polymer adopts a compact, nonfractal conformation. $R^2(s)$ shows a plateau behavior at large s, with an exponent $\nu \sim 0$, also often observed in FISH data [9, 10] (Fig. 1.6A). The shape of $P_c(s)$ as a function of genomic distance, s, also reflects the three regimes described above (Fig. 1.1E). Similarly to $R^2(s)$, $P_c(s)$ has a power-law behavior with $P_c(s) \sim 1s^{\alpha}$, where the exponent α is also dependent on c_m (Fig. 1.1E). When $c_m < C_{tr}$, $\alpha = 2.1$, which is the signature of the randomly folded (open) polymer (SAW model) [33]. For c_m around C_{tr}, the scaling exponent changes in a range encompassing $\alpha = 1.08$ found in Hi-C data [12]; we find $\alpha = 1.5$ at the transition Θ point. Distant loci are more likely to contact each other than in the free open polymer. When $c_m > C_{tr}$, the polymer shrinks into a compact mass manifested by a plateau of $P_c(s)$ at large s, where the exponent becomes $\alpha = 0.0$.

1.1.3 Comparison of the SBS and Other Models Against Experimental Data

Before delving further into implications of the SBS model, we compare its predictions against experimental data and those of other models. The FG model represents the chromatin fiber as a noninteracting (free) polymer chain in a specific transient state. It was proposed in 1988 in the polymer physics literature as a knot-free state [15], and used recently to explain the behavior of $P_c(s)$ from Hi-C data and to propose that chromatin is organized as fractal globules [12, 36]. The FG model seems attractive because it proposes that chromatin is found in a specific, unique fractal state that resembles the 1-Mb chromatin domains suggested earlier [14]. Although the FG model only considers random chromatin interactions and not binder-mediated contacts as identified experimentally, it provides a $P_c(s)$ with an exponent α of approximately 1, which is very close to the value of α estimated from some Hi-C average data [12]. Thus, the FG model depicts chromatin as if it were all in a single conformational state. Importantly, it also predicts that $R^2(s)$ grows indefinitely with s with an exponent ν of approximately 0.33. However, although ν around 0.33 can be observed at some specific loci and cell types, it is not a general value found across most experimental datasets where a plateau in $R^2(s)$ is often observed (Fig. 1.6A). Furthermore, the FG state is only achieved using highly specific simulation conditions. For instance, the polymer must be initially forced into a highly compacted state without knots, before being released and becoming fully unfolded. The time window during which the polymer behaves as a FG only exists fleetingly, and the polymer converges to a different equilibrium state. This time window would become vanishingly small in the presence of key nuclear factors, such as DNA topoisomerases [36]. To compare predictions from SBS and FG models, we first considered available

FISH data on $R^2(s)$ from different chromosomes and systems: on chromosome 12 in pro-B cells ($0-3$ Mb; Fig. 1.2A) [10], and on chromosome 11 in primary fibroblasts ($0-80$ Mb; Fig. 1.8) [36]. The SBS model in the closed polymer state correctly fits all sets of FISH data, representing the early increase and plateau in $R^2(s)$ at shorter and longer genomic distances, respectively. In contrast, the FG predicts that $R^2(s)$ grows indefinitely with s ($\nu \sim 0.33$). Thus, the FG model accounts only for the early increase in $R^2(s)$ (Fig. 1.2A and Fig. 1.8), but fails to capture the leveling off at longer s. Interestingly, the $R^2(s)$ plateau across these two cell systems arises at different s, reflecting biological complexity (unless related to methodological differences). We next investigated the generality of the value of α around 1.08 derived after averaging $P_c(s)$ across all chromosomes in the human female lymphoblastoid cell line (GM06990) [12]. Using the published Hi-C data, we calculated $P_c(s)$ for different chromosomes separately (Fig. 1.2B and Figs. 1.9A and 1.10A). To investigate the effects of chromatin compaction, we chose to compare chromosomes 19 (gene dense with high gene expression) and X (one copy is silent in this female cell line). We show that chromosomes 19 and X deviate from the average behavior in the 0.5-7 Mb region, with α exponents ranging from 0.93 and 1.30, respectively. This is consistent with their average open and closed states, compared to chromosomes 11 and 12, which have α of approximately 1.08. We observed a similar deviation from the average $P_c(s)$ in a different female lymphoblastoid cell line (GM12878) analyzed by either Hi-C or tethered conformation capture (TCC) [16], and in IMR90 cells characterized by Hi-C [37] (Fig. 1.9B-D). Analogous comparisons between chromosomes 18 (gene poor) and 19 (gene rich) yield similar deviations from the average behavior, with chromosome 18 having larger α than chromosome 19 (Fig. 1.10). Surprisingly, analyses of Hi-C data from the human embryonic stem cells [37] (H1-hESC) showed a striking deviation from the lymphoblastoid cells analyzed above. In H1-hESC, averaged Hi-C contact probabilities for all individual chromosomes analyzed resulted in a higher α of approximately 1.6 (Fig. 1.2C). This result agrees with previous findings that stem cell chromatin tends to assume more open conformations than in other cell types [38]. Direct comparisons of genome-wide $P_c(s)$ reveal different exponents α across the cell lines studied (Fig. 1.13 and Fig. 1.9F). We stress that calculations of α for whole genomes or chromosomes reflect average chromatin folding behaviors that disregard the variety of conformations known to exist at specific loci. To illustrate this concept, we investigated whether the Hi-C-derived values of α could in principle be obtained by simple averaging over regions of open and closed chromatin, even in the absence of fractal folding states. Thus, we considered a mixture of SBS model systems containing a proportion of open and compact polymers (p and $1-p$, with $\alpha 2.1$ and 0.0, respectively; Fig. 1.1F). The average $P_c(s)$ of such mixtures has an exponent α that depend on the proportion p (Fig. 1.2D). Strikingly, $\alpha = 1.08$ can be found for p of approximately 0.60, in a range of s that spans one order of magnitude, as observed in Hi-C data. When

the fraction of open polymers is decreased to $p = 0.45$, $\alpha = 0.93$ in the same s range, a value close to the one found for chromosome X in the female cell line GM06990 [12]. Conversely, $p = 0.80$ gives $\alpha = 1.3$, as it does for chromosome 19 in the same cell line. This analysis illustrates the lack of power of the average α exponent alone to elucidate chromatin architecture. Taken together, our results strongly argue that an average α of approximately 1.08 does not describe a general principle of chromatin behavior, and show that the FG model is not a general description of chromatin folding principles. In contrast, the SBS model has the power to explain the whole range of α exponents identified experimentally (Fig. 1.13 and Figs. 1.6A, 1.7, 1.9, and 1.10). Nevertheless, the FG model may help explain specific transitional states for some genomic regions, as, for instance, rapid chromatin decondensation and chromatin looping out of chromosome territories during gene activation [39–43].

1.1.4 The SBS Model Reproduces the Organization of Chromatin in Topological Domains

Chromatin domains or globules have been hypothesized as a basic unit of chromatin organization, based on the appearance of chromatin seen by electron microscopy, the size of chromatin loops, and evidence for chromatin associations derived by nuclear structures such as clustering of replicons in replication factories [3, 14, 44]. Recent analyses of 3C-based studies are also consistent with the existence of approximately 1-Mb domains [37, 45]. To explore the formation of chromatin globules in the SBS model, we modeled a polymer containing two kinds of binding site (red and green), segregated in two separate halves of the polymer length, each with specific affinity to one kind of binder (red and green, respectively; Fig. 1.3A). In these conditions, the SBS model promptly produces separate domains, as demonstrated on single simulations (Fig. 1.3B) or by the average matrix of interactions (Fig. 1.3C). As expected, the $P_c(s)$ of such a polymer has two different regions (Fig. 1.3D). At s shorter than the length of each domain, $P_c(s)$ has a lower value of α, corresponding to a closed (globular) state. At larger s, it has a higher α, corresponding to an open state. The mean-square distance, $R^2(s)$, also has two regions, with an increase followed by a plateau as s increases (Fig. 1.11). As a final example of the power of the SBS model to simulate known chromatin behaviors, we investigated whether changes in binding site affinity upon domain formation could induce chromatin looping out from domains [39–43] (Fig. 1.12). We let the two globules reach equilibrium (as for Fig. 1.3), but subsequently changed the state of three (out of 11) contiguous sites from binding the red binders to no longer having affinity to binders (becoming blue). Red and green domains remained (red domains now of smaller dimension), but the polymer segment containing blue binding sites looped out from the

red domain (Fig. 1.12). Analogously, associations between different domains can be explained by the action of common binders (SI Text). Chromatin domains and looping can therefore be studied with the SBS model in conditions of full segregation or of partial mixing, depending on the binding site geography and binder properties. With appropriate experimental data of chromatin associations and epigenetic mapping, the SBS model has the power to describe principles that drive chromosome folding and the dynamic changes occurring during differentiation and in disease.

1.1.5 The SBS Model Reproduces the Dynamic Folding Behaviors of Chromatin

As a final test to the power of the SBS model, we investigated the kurtosis $K = \langle R^4(s)\rangle / \langle R(s)^2\rangle^2$, which is the ratio of the fourth and second moment of the spatial distance. K is a dimensionless quantity that allows for direct comparisons between polymer modeling results and experimental data [33]. In principle, K carries information about genomic locus-to-locus and cell-to-cell variations. Interestingly, K has been measured experimentally by FISH [9, 46, 47] and found to vary between 1.5 and 4.4, depending on the locus and cell type studied (Fig. 1.4A). Importantly, the SBS model produces a range of K values that span the same range of experimental K. In fact, K depends on c_m (Fig. 1.4B). K is approximately 1.5 at both low and high concentrations of binders (i.e., open and closed chromatin, respectively), but in the region around the threshold concentration K increases from 1.5 up to 5, encompassing all previously reported experimental measurements. To date, no other chromatin polymer model produces such a variety of behaviors and correspondence with experimental results. The experimental observations [9, 46, 47] of loci where $K = 1.5$ (the value predicted by the SBS model for open and compact states; see Fig. 1.4A) and of loci with values between 1.5 and 5 (corresponding to the transition region between open and closed states) suggest that, indeed, chromatin exists in a complex mixture that includes fractal as well as open and closed states. Measurements of K from previous models produce a constant value of K. The randomly folded polymer (SAW) model produces $K = 1.5$ [33]. Previous polymer models that consider the effects of both specified chromatin loops at fixed lengths of $120k$b and 3 Mb [11, 48], and the possibility of chromatin loops of all sizes [49], also produce constant K values. Furthermore, although these models explain the plateau of the mean-square spatial distance, $R^2(s)$, at large s, they do not reproduce the behavior of the contact probability, $P_c(s)$, observed in Hi-C data, as does the SBS model. More recently, the possibility of transient interactions across a polymer has been studied in the dynamic loop (DL) model, which, to mimic the effects of bridging molecules, assigns an attachment probability to genomic regions that randomly meet in space [47]. In

agreement with the SBS model, the DL model can also reproduce the Hi-C exponent of $P_c(s)$. K has not been calculated for the FG model, but it will by definition give a constant K.

1.2 Materials and Methods

1.2.1 Model and its parameters

In the "strings and binders switch" (SBS) model, a chromatin filament is represented as a self-avoiding polymer chain [19]. Here, the chain is made of $n = 512$ spherical sites, each s_0 bases long (total length $L = n \cdot s_0$). Because of their scaling properties, modeling of shorter or longer polymers displays similar behaviors [18, 19, 26]. The polymer has a fraction, f, of binding sites (here, we consider mainly the case $f = 0.5$; see ref. [19] for the general case) and interacts with a concentration, c_m , of diffusing molecules (binders), which have an affinity E_X for those polymer sites. Real transcription factor binding energies range from approximately $2k_BT$ for nonspecific binding sites to approximately $20k_BT$ for specific ones (k_B is the Boltzmann constant and T is the temperature in kelvin). Here, for simplicity, we set $E_X = 2k_BT$ for all sites (see ref. [19] for the general case). In real situations, the number and location of binding sites depend on the specific locus considered. For definiteness, we consider here the simplified case where they are evenly distributed along the polymer chain. The system is investigated by Metropolis Monte Carlo (MC) simulations, and for computational purpose the polymer inhabits a cubic lattice with lattice spacing, d_0, equal to the linear length of a polymer site. Diffusing molecules and polymer sites randomly move from one to a nearest-neighbor vertex on the lattice, with single occupancy. Polymers obey a nonbreaking constraint: two proximal sites can sit either on next or nearest-next neighboring lattice sites. Chemical interactions are only permitted between nearest-neighbor particles. Our binding molecules have a binding multiplicity of six, after considering the characteristics of known chromatin organizers, such as CCCTC-binding factor (CTCF), capable of multiple binding, or other organizers, such as transcription factories. MC averages are over up to 10^3 runs, each run being up to $10^1$1-step long. The linear length of a polymer site, d_0, can be roughly estimated as: d_0 is approximately $(s_0/G)^{1/3}D_0$, where D_0 is the nucleus diameter and G is the genome content. In mammals, we have approximately $G = 6G$b and D_0 of approximately 5μm. Thus, for a polymer of $L = 10M$b, we have $s_0 = L/n$ of approximately $20k$b, and we derive d_0 of approximately 0.1μm. The fraction, c, of molecules per lattice site is related to their molar concentration, $c_m c \sim c_m d_0^3 N_A$, where N_A is the Avogadro number. We explored the system behavior as a function of

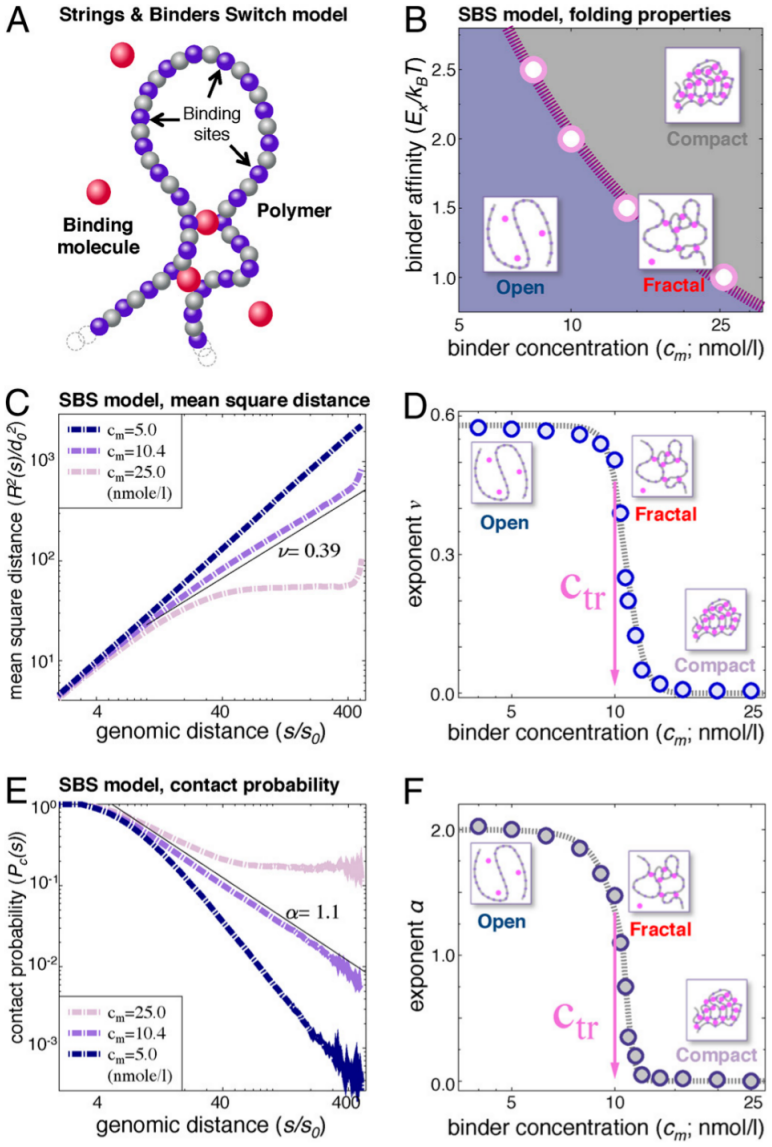

FIGURE 1.1: The emerging stable states of the SBS model and the mechanisms of its self-organization. (A) Schematic representation of the SBS model. A chromatin filament is represented by a SAW polymer comprising n beads randomly floating within an assigned volume. A fraction, f, of beads (binding sites) can interact with Brownian molecules (magenta spheres; binders) with concentration c_m and binding site affinity E_X. In this example, $f = 0.5$ for an equal number of blue (binding) and grey (non-binding) sites. Molecules bind more than one polymer site, allowing for loop formation. (B) Three classes of states exist. The phase diagram illustrates the conformational state of the system as a function of two main control parameters, c_m and E_X . The system is in an open randomly folded conformation below the transition line, $C_tr(E_X)$ (dashed curve), it folds in a compact conformation above it, and it takes a different fractal structure at the transition point. (C) The polymer mean-square distance. $R^2(s)$ is the mean-square distance (in units of the bead linear length d_0) of two polymer sites having a contour distance s. R^2 is shown as a function of s for three values of the binder concentration, $c_m = 5$, 10.4, and $25n\text{mol}/L$, corresponding to below, around, and above the transition point (here, $E_X = 2k_BT$). At large s, $R^2(s)$ has a power-law behavior $R^2(s) \sim s^{2v}$; at $c_m = 10.4n\text{mol}/L$ we find $\nu \sim 0.39$. For $ss_0 > 400$, finite size effects are seen. (D) The power law exponent of $R^2(s)$ has three regimes. The exponent, ν, has a sigmoid behavior as a function of c_m , corresponding to different system states, with $\nu \sim 0.58$ for $c_m < C_tr$; $\nu \sim 0.5$ at $C_tr = 10.0n\text{mol}/L$; and $\nu \sim 0.0$ at $c_m > C_tr$. (E) Site contact probability. $P_c(s)$ is the contact probability of two sites with contour distance s along the polymer chain. It is plotted for $c_m = 5$, 10.4, and $25n\text{mol}/L$. At large s, a power law is found: $P_c(s) \sim 1s^{\alpha}$. We find $\alpha = 1.1$ at $c_m = 10.4n\text{mol}/L$. (F) Power law behavior of $P_c(s)$. The $P_c(s)$ exponent α expressed as a function of c_m also displays three regimes: below, around, and above C_tr.

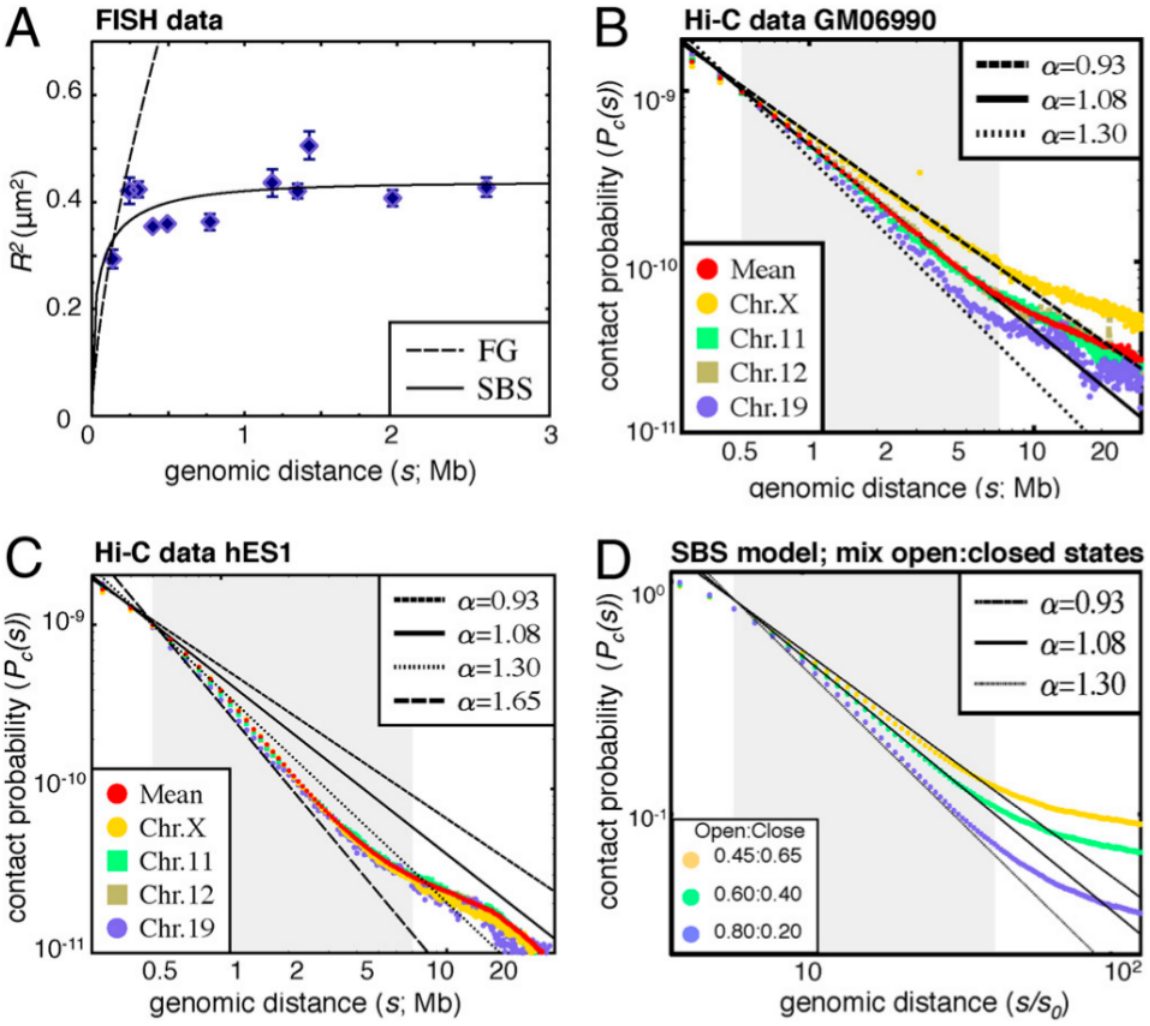

FIGURE 1.2: The SBS model explains the range of experimental chromatin folding behaviors. (A) Mean-square distance of subchromosomal regions from FISH data. Mean-square distance, $R^2(s)$, from FISH data in pro-B cells chromosome 12, spanning 3 Mb [10]. Superimposed dashed line indicates behavior predicted by the FG model; continuous line indicates behavior predicted by the SBS model in the compact state. (B-D) Contact probability from Hi-C data and SBS model. (B) Contact probability, $P_c(s)$, was calculated separately for different chromosomes from published Hi-C dataset in human lymphoblastoid cell line GM06990 [12]. Chromosomes 11 and 12 follow the average behavior reported [12] in the 0.5-7 Mb region (shaded in grey), with exponent α of approximately 1.08. Chromosomes X and 19 deviate from the average, with α exponents of approximately 0.93 to approximately 1.30, respectively. In a given system, different chromosomes can have different exponents. (C) $P_c(s)$ was calculated for different chromosomes from published Hi-C dataset in human embryonic stem cell line H1-hESC [37]. All chromosomes deviate from exponent α of approximately 1.08 in the 0.5-7 Mb region (shaded in grey), and have an exponent α of approximately 1.65, characteristic of open chromatin within the SBS interpretation. Different systems can have different exponents. (D) Mixtures of open and compact SBS polymers can model average $P_c(s)$. Average $P_c(s)$ is shown for mixtures of open and compact polymers in the SBS model (where α2.1 and 0.0, respectively). In each mixture, p and $1-p$ are the fractions of open and compact polymers, respectively. $P_c(s)$ and α depend on p. For p of approximately 60%, $\alpha = 1.08$ is found in a range of s about one order of magnitude long, as in Hi-C data. Simply changing the fraction of open chromatin can recover the entire range of Hi-C exponents of B and C.

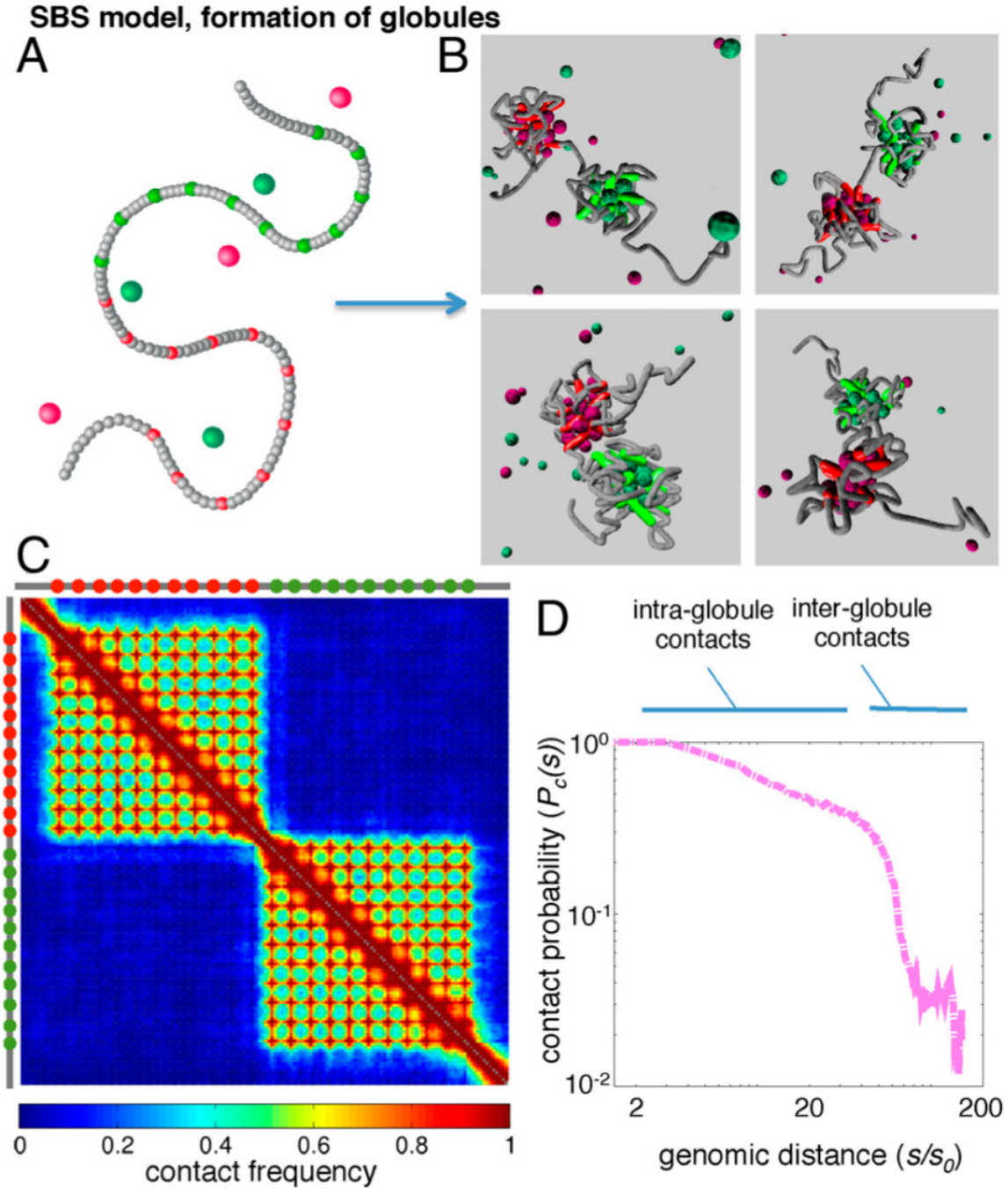

FIGURE 1.3: The SBS model captures the globular conformation of chromatin. (A) Schematic representation of the polymer system used to study formation of chromatin globules or domains. (B) Snapshots of chromatin domains formed after MC simulations (Appendix B) of the SBS polymer model represented in A. (C) The steady-state contact matrix shows two separate chromatin do- mains. (D) The average $P_c(s)$ shows two regimes: closed chromatin, at shorter s because of formation of globules; and open chromatin, at larger genomic regions because of the absence of interactions between the two domains.

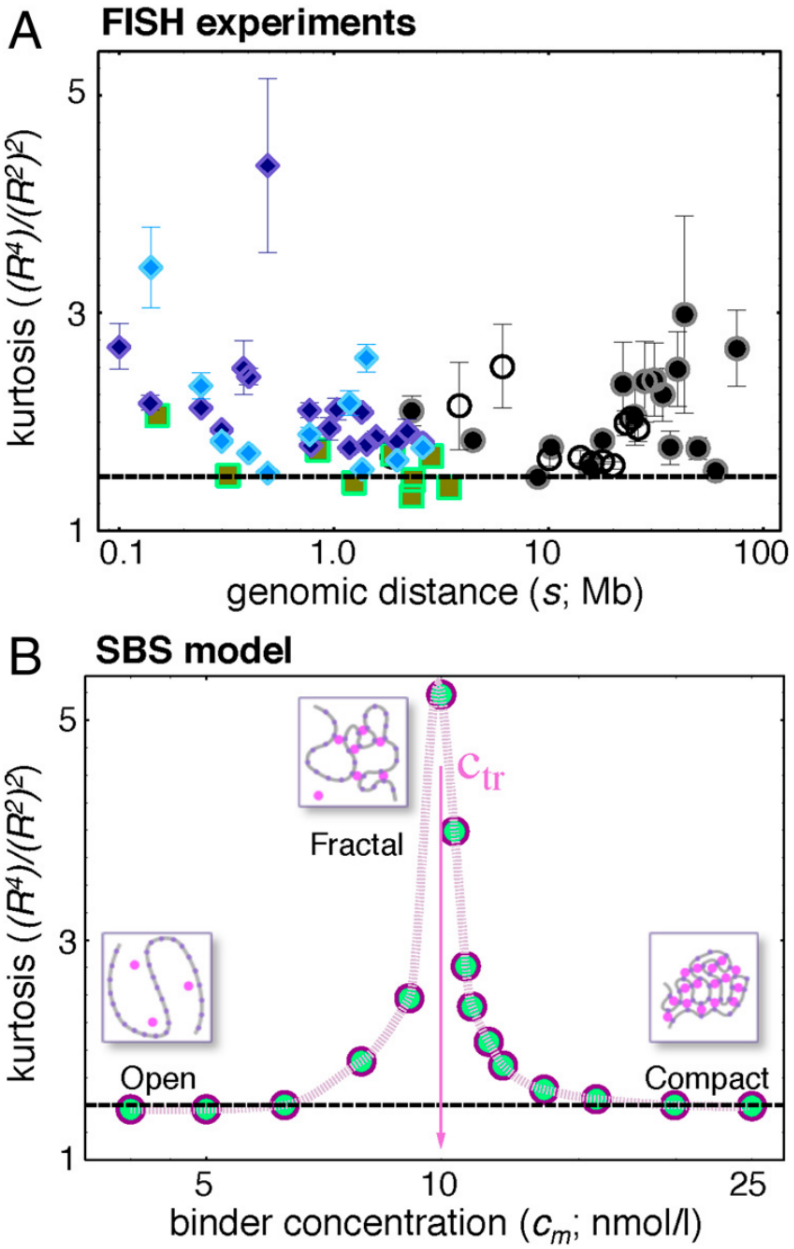

FIGURE 1.4: The SBS model captures the full range of values of the distance kurtosis observed in FISH data. (A) The ratio of the fourth and second moment of the distance R between loci at the genomic distance s [i.e., the kurtosis; $K = \langle R^4(s) \rangle / \langle R(s)^2 \rangle^2$] is plotted as a function of s. It provides a measure of the relative amplitude of fluctuations of the polymer conformations. $K = 1.50$ when $R^2(s)$ is randomly distributed as a self-avoiding polymer (horizontal dashed line). Experimental K values depart from 1.50; K values were first analyzed in ref. [47], and originate from human fibroblast chromosome 1 ridges or whole chromosomes 1 or 11 [9] (squares, open circles, and filled circles, respectively), and from pre/pro-B or pro-B cell murine immunoglobulin heavy chain locus [10] (light- or dark-blue diamonds, respectively). (B) The kurtosis measured in the SBS model is plotted as a function of c_m. K is close to 1.5 at low and high concentrations of binding molecules (open and closed chromatin). Around the binder threshold concentration, K exhibits a peak with values up to approximately 5. The range of values of K measured experimentally (A) matches the range found within the SBS model. It emerges that, beyond open and compact states, chromatin loci are likely to

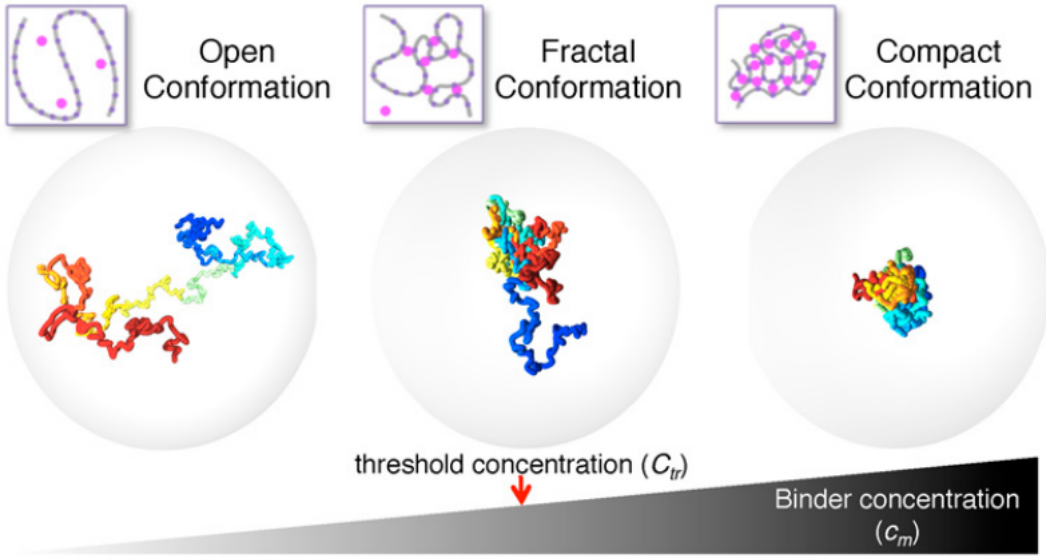

FIGURE 1.5: Overview of the system states and their transitions. Representation of three classes of stable conformational states of the SBS polymer chain shown in Fig. 1: (Left) the open random coil ($c_m = 5n\text{mol}/L$; $c_m < C_tr$), (Center) the transition-point fractal ($c_m = 10n\text{mol}/L$; c_m around C_tr), and (Right) the compact globule state ($c_m = 25n\text{mol}/L$; $c_m > C_tr$). The polymer conformations were obtained from MC simulations of the SBS model (Appendix B). For clarity, the polymer binding molecules are not shown and the surrounding transparent sphere represents the nucleus. Polymer and sphere sizes are proportional to the size of mammalian chromosomes and nuclei, respectively. Switch-like conformational changes occur, regulated by increasing c_m or E_X above precise threshold values marking thermodynamic phase transitions.

c_m (given $E_X = 2k_BT$) and analogous findings are encountered by changing E_X (given c_m).

1.3 Approach to stationarity

MC Metropolis simulations at stationarity correctly describe the general equilibrium state of a system [50]; here, we focus on such equilibrium states. In the current prevailing interpretation (in systems dominated by Brownian motion, ref. [50]), a MC Metropolis dynamic is also accepted to describe correctly its general long-term evolution. We therefore tested whether the orders of magnitude of the time scales to approach equilibrium, predicted by our MC dynamics, fall within the range expected biologically [51]. The MC time unit corresponds to a time $t_0 = 1/r_0$ [50], which is related to the polymer diffusion constant D and to the lattice spacing constant $d_0 : D = S^2/(d_0^2/4t_0)$, where 2 is the mean-square displacement of the polymer center of mass per unit of MC time. We measure S^2 and impose that D is of the order of magnitude of the measured diffusion constant of mammalian DNA loci ($D = 1mm^2/h$) [52] to derive t_0. In this

way, we can map the MC time steps into real time. The folding time predicted by MC is of the order of magnitude of fraction of hours, which, interestingly, is in the biological expected range.

1.4 Globule formation and topological domains

With the SBS model we also explored the mechanisms of formation of distinctly folded polymer domains, as recently discovered in chromatin studies [37, 45]. To this end, we considered a variant of our model (Fig. 1.3A) where there are two kinds of equally spaced binding sites ($f = 1/6$) along the polymer chain ($n = 152$ beads long). The first type of binding site, shown in red, is located in the first half of the polymer chain, whereas the second type, shown in green, lies in the second half. Importantly, the two types of site interact with a distinct molecular binder each (here, each present in a concentration $c_m = 25n\text{mol}/L$, and with $E_X = 4k_BT$), red sites with red molecules and green sites with green molecules (here, molecules do not interact with each other). We explored such a system and now discuss the case where both the red and green binders are in a concentration high enough to drive red and green polymer sites in their compact folded states (i.e., $c_m = 25n\text{mol}/L$, and $E_X = 4k_BT$). As stationarity is approached, two globules spontaneously appear along the polymer, one composed of red and the other of green sites, as reported in the pictures of Fig. 1.3B. For such a system we also computed the contact matrix (i.e., the probability that any two sites along the polymer are in contact). This is shown in Fig. 1.3C: Two distinct domains appear in the contact matrix, with strong intradomain and much weaker interdomain interactions, simply explained here by the action of the two distinct, specific binding molecules. Interestingly, the formation of such separated globules closely resembles the topological domains observed in chromatin by recent genome-wide chromosome conformation capture (Hi-C) and Carbon-Copy Chromosome Conformation Capture (5C) studies [37, 45]. The presence of the two domains is also manifested in the average contact probability, $P_c(s)$, plotted in Fig. 1.3D, which shows a shoulder and a change of slope at genomic distances, s, corresponding to the crossover from one to the other polymer region (roughly around $s = n/2$). Analogously, after an early increase with s the mean-square distance, $R^2(s)$, develops a plateau, corresponding to the compact folded structure of the polymer domains, as seen in Fig. 1.11. In Fig. 1.12 A and B, we also show the equilibrium distance distribution of four specific sites along the polymer chain: sites A, B, C, and D. It appears that A and B have a short distance and narrow distribution because they belong to the same domain (analogously, C and D), whereas the distance distribution, for instance, of A and C (or A and D) is broader and centered around much higher values because A belongs to a distinct domain with relation to C (and D).

1.5 Associating domains

Associations between different domains can be generated if they share binding sites of a given kind. In the same way that associations of sites within distinct topological domains (red/green) are determined (e.g., by the presence of shared binding molecules), different domains can associate with each other when some of their binding sites have common binding molecules (e.g., yellow). Topological associated domains (TADs) can be produced in this way, as recently seen in Hi-C and 5C experiments [37, 45].

1.6 Looping out of specific sites from their domains

To investigate the mechanisms of looping out of a given polymer site and explore domain formation, we also took advantage of the two-binding site polymer system considered in Fig. 1.3. For this purpose, we let the two (red and green) domains form before changing the three red binding sites centered on site A to be in an inert state (i.e., we impose that they are no longer able to interact with diffusing binding molecules; see Fig. 1.12 C-F, where the sites that changed state are highlighted in blue). The system starts from a conformation with the two compact domains (red/green) discussed above and is allowed to approach the new equilibrium state induced by the presence of the new blue sites, including site A, which are inert. A conformation of the system at stationarity and the new contact matrix are shown in Fig. 1.12E; the signature of the red and green domains is still clear, but now the region around site A has lost contact with its former domain (compare to the contact matrix in Fig. 1.3C of the two-domain state previously discussed). Consistently, the distance distribution of A from B opens up to become similar to the distribution of distances from C or D (compare Fig. 1.12F and B). This example illustrates how looping out of a site from its domain can easily be modeled within the SBS model.

A

Cellular material	Technique	Genomic region analysed (resolution)	Mean square spatial distance, $R^2(s)$, as a function of genomic separation, s	Contact probability, $P_c(s)$, as a function of genomic separation, s	Ref.
Human female fibroblast cells	3D-FISH	Chr1q, 27Mb (~1Mb) and Chr11q, 75Mb (~3Mb)	s = 0.4-2Mb, $R^2(s)$ increases (ν = 0.33) s >10Mb, $R^2(s)$ reaches a plateau (ν = 0)	N/A	9
Mouse pre-pro-B cells (E2A$^{-/-}$) and pro-B cells (RAG$^{-/-}$)	3D-FISH	Igh locus, 3Mb (~300kb)	s <0.5Mb, $R^2(s)$ increases (ν = 0.25 for pre-pro-B cells, ν = 0.1 for pro-B cells) s >0.5Mb, $R^2(s)$ reaches a plateau (ν = 0; both cell types)	N/A	10
Mouse NIH-3T3 fibroblasts	3D-FISH	Chr14, 4.3Mb (200kb)	s <3.5Mb, $R^2(s)$ increases ($\nu \sim 0.5$) s >3.5, $R^2(s)$ may plateau	N/A	11
Human male fibroblast cells	2D-FISH	Chr4 (10Mb)	s <50Mb, $R^2(s)$ increases ($\nu \sim 0.5$) s >50Mb, $R^2(s)$ increases ($\nu \sim 0.3$)	N/A	12
Human lymphoblastoid cell line	Hi-C	Genome-wide (1Mb)	N/A	s = 0.5-7Mb, $P_c(s)$ decreases approximately as a power law, with exponent $\alpha \approx 1.08$	13
Drosophila embryos	Simplified Hi-C	Genome-wide and Repressive epigenetic classes	N/A	Genome wide: contact frequencies decrease approximately as a power law, with exponent $\alpha \approx 0.85$ Repressive epigenetic classes: contact frequencies decrease approximately as a power law, with exponent $\alpha \approx 0.70$	14

B

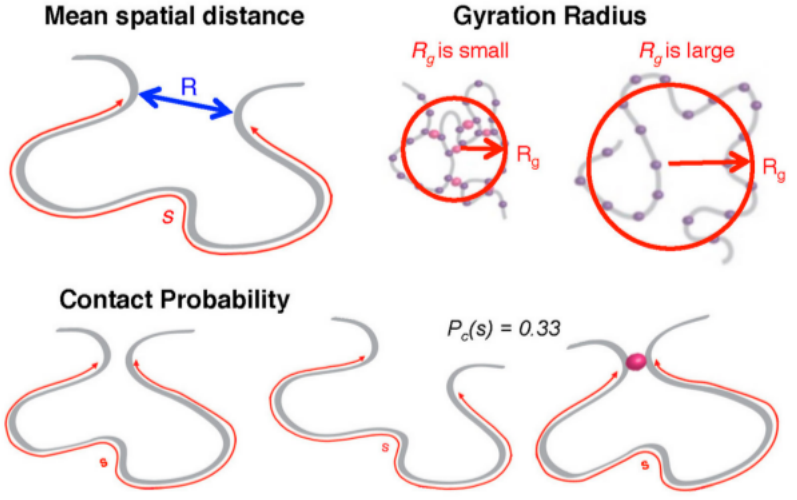

FIGURE 1.6: Chromatin folding exhibits a variety of folding behaviors that are locus and cell-type specific. (A) FISH and Hi-C data (Appendix A). The table summarizes published data on the behavior of the mean-square distance, $R^2(s)$, and contact probability, $P_c(s)$, as a function of genomic distance, s, found in FISH and Hi-C experiments (Appendix A). The reference numbers given in this table refer to the list of the main text. (B) Summary of parameters commonly used to quantify chromatin folding behaviors. The figure illustrates the definition of linear genomic distance (s), spatial distance (R), contact probability of two loci [$P_c(s)$], and radius of gyration of the polymer (R_g) representing the radius of the average sphere enclosing the polymer.

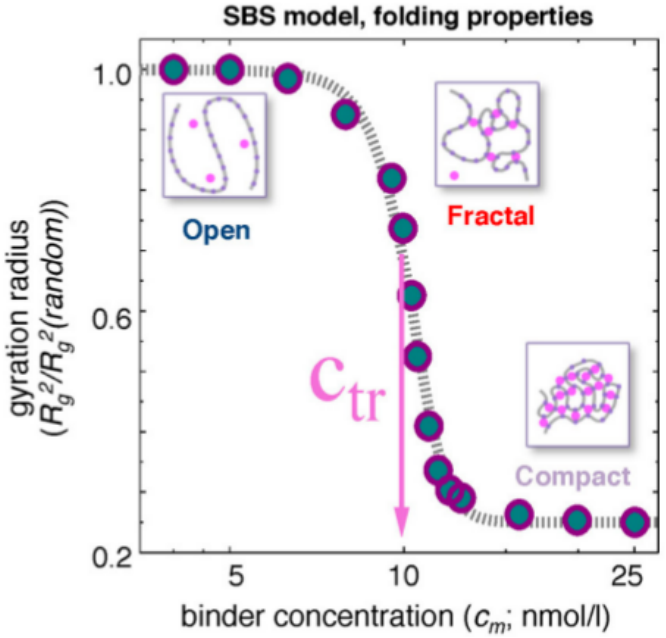

FIGURE 1.7: The polymer conformation displays switch-like responses to changes in binder concentration. The polymer equilibrium-squared gyration radius, R_g^2 , is plotted as a function of binder concentration, c_m (for $E_X = 2k_BT$). Rg is normalized by the gyration radius of a random SAW chain of equal contour length N. R_g^2 has a step-like behavior. Below a transition value of C_{tr} (C_{tr} is approximately $10n\text{mol}/L$ for $E_X = 2k_BT$), R_g^2 is indistinguishable from the value found for the random SAW polymer. Above C_{tr} , R_g^2 takes a value corresponding to a chain folded in a compact spherical globule. (Insets) Schematic drawings of the polymer conformation in the different states.

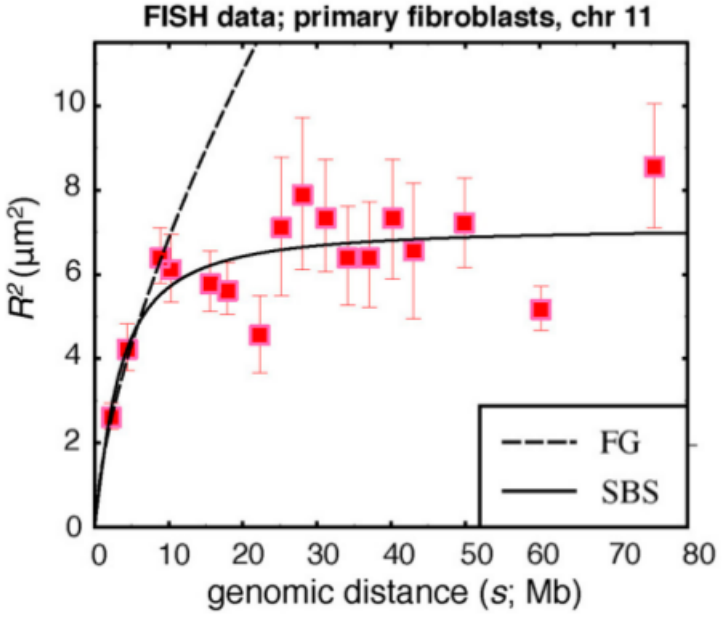

FIGURE 1.8: The mean-square distance of subchromosomal regions from FISH data (Appendix A). FISH data on the mean-square distance, $R^2(s)$, in primary fibroblast chromosome 11, spanning $80M$b [45]. The superimposed dashed line is the behavior predicted by the FG model, and the continuous line the behavior predicted by the SBS model in the compact state.

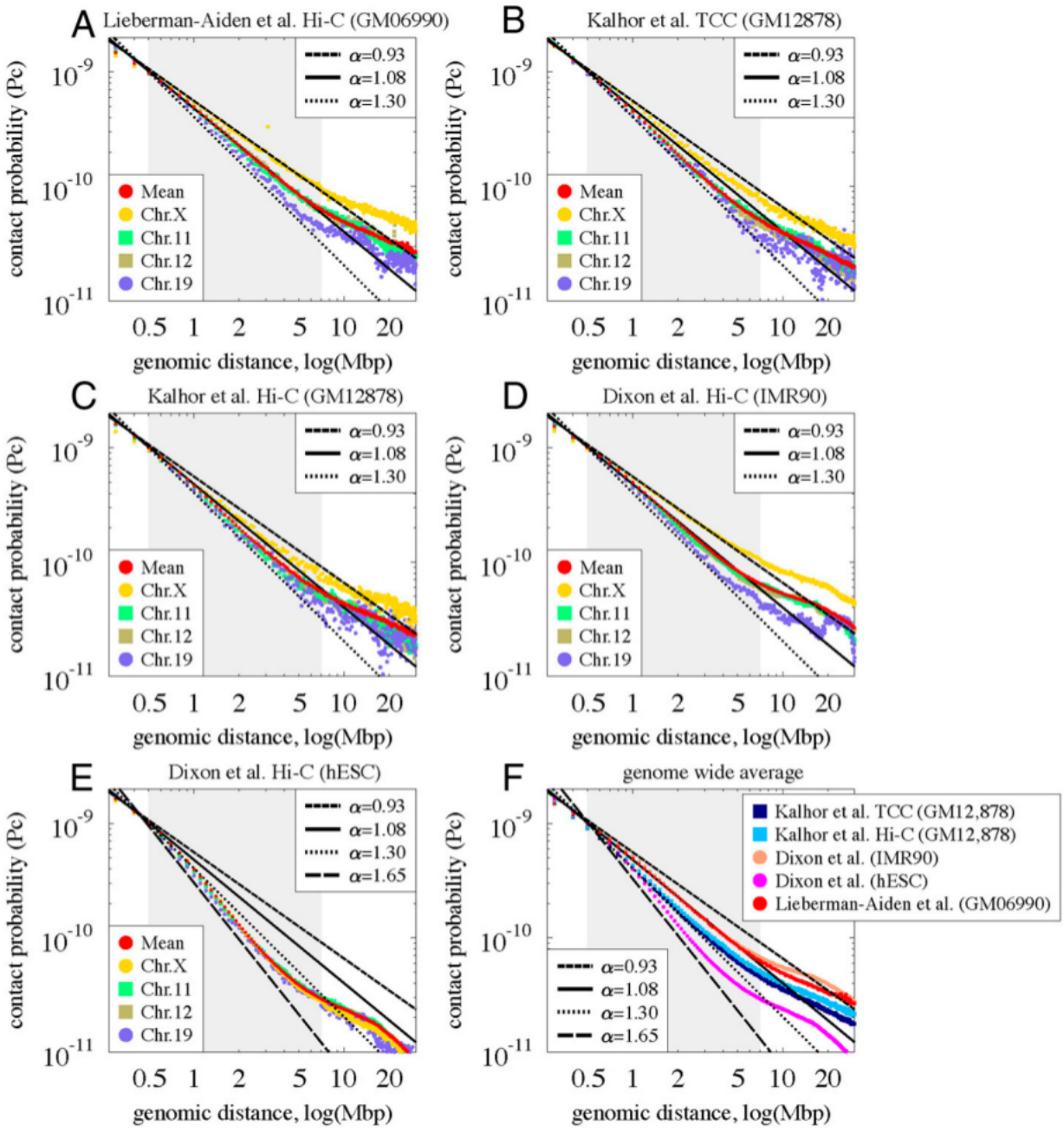

FIGURE 1.9: Contact probability across Hi-C and TCC experiments. Hi-C and TCC contact probabilities for different chromosomes in different cell types (Appendix A). (A) $P_c(s)$ is shown for different chromosomes separately from the published Hi-C dataset for GM06990 cell line [12]. Chromosomes 11 and 12 follow the average behavior envisaged by Lieberman-Aiden et al. [12] in the $0.5 - 7$ Mb region (shaded in grey), having an exponent α of approximately 1.08. Chromosomes 19 and X deviate from the average, with α exponents ranging from approximately 0.93 to approximately 1.30. (B-D) $P_c(s)$ analyses from Hi-C and TCC datasets from GM12878 cell line [16] or IMR90 [37] show a behavior similar to the GM06990 cell line. (E) $P_c(s)$ analyses from Hi-C for H1-hESC [37]: A distinct behavior is seen for H1-hESC, where all chromosomes appear to have an exponent close to 1.6, which corresponds to a more open conformation in the SBS model (see main text). The same results are seen in H1-hESC whether or not the sex chromosomes are included in the analyses. (F) Comparison between the genome average $P_c(s)$ observed in the different cell lines considered, showing different exponents and behaviors.

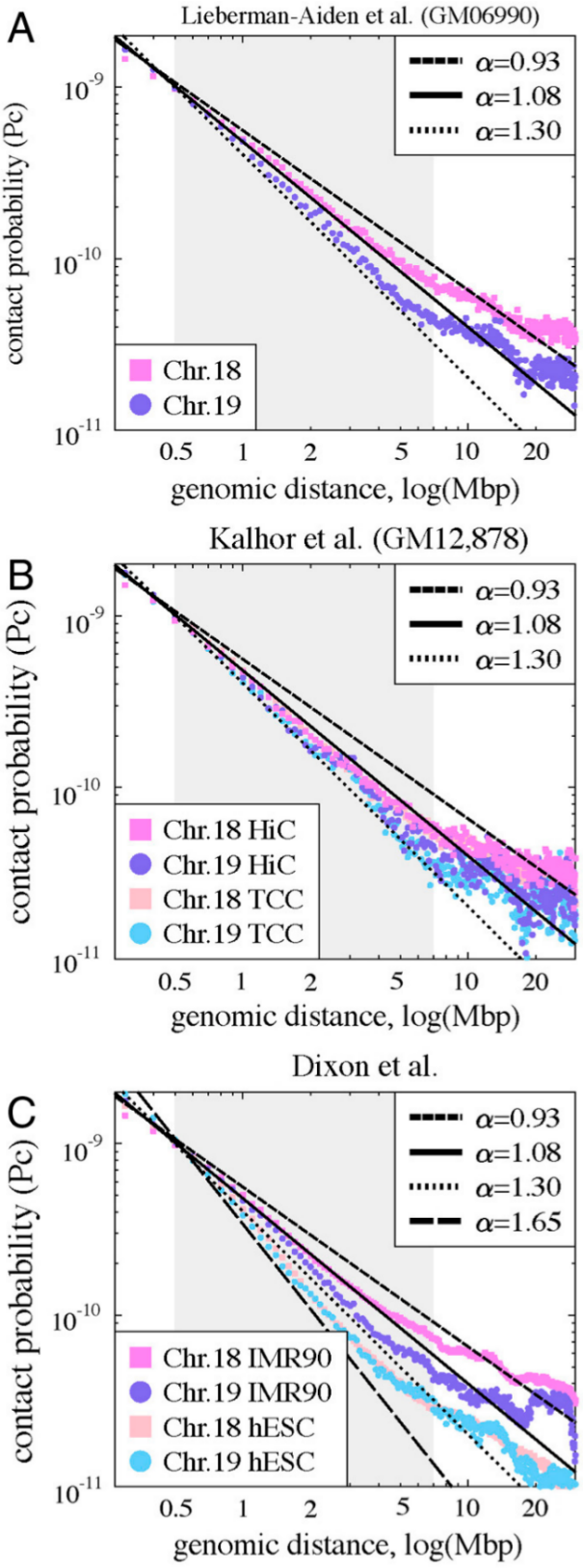

FIGURE 1.10: Comparison of the contact probability of chromosome 18 and 19 within individual experiments of Fig. 1.9.

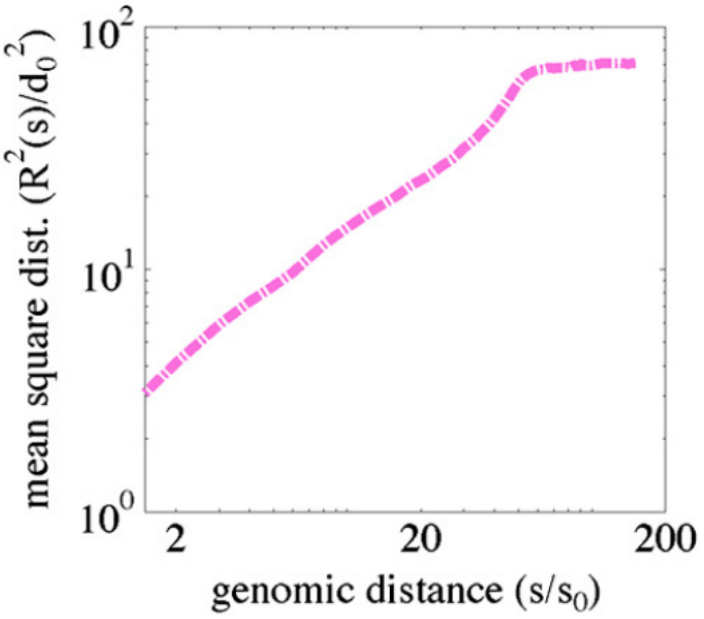

FIGURE 1.11: Mean-square distance, R^2, between two polymer sites having a genomic distance, s, in the two-domain state of the model pictured in Fig. 1.3A.

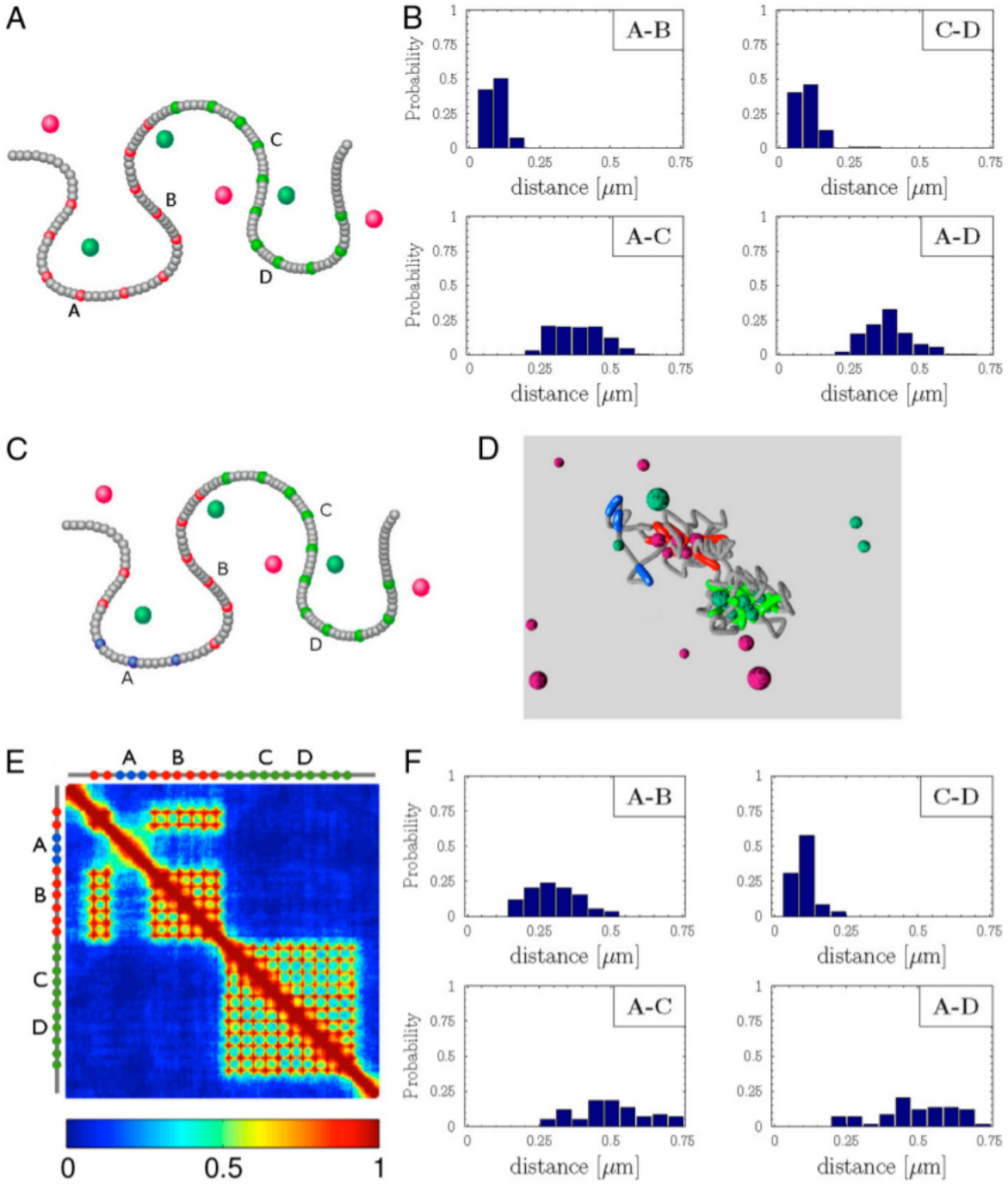

FIGURE 1.12: Modeling chromatin looping out of domains. (A) Scheme of the polymer of Fig. 1.3 with the position of sites A, B, C, and D illustrated. (B) Distance distribution of the four specific sites A to D at equilibrium. Note that A and B, and C and D belong to red and green domains, respectively. (C) Schematic representation of the polymer considered to study looping of site A from the red domain. Once the red/green polymer represented in A reaches equilibrium, sites represented in blue are modified to loose interaction with red moleculesi.e., they change from red (with affinity to red binders) to blue (without affinity to red or green particles). (D) Picture of a looped-out configuration of site A and neighbor sites from MC simulations. (E) The steady-state contact matrix of the polymer system showing that site A is no longer inside the red domain. Color scale from blue to red indicates contact frequencies. (F) Distance distribution of sites A-D, when site A becomes inert (blue). Site A loops out of its domain: In comparison to distances represented in B, distances between sites A and B are now longer, whereas C and D (for comparison) keep their relative positions into their green domains.

	GM06990 HiC*	GM12878 HiC†	GM12878 TCC†	IMR90 HiC‡	H1–hESC HiC‡
Mean	1.1	1.2	1.3	1.1	1.6
Chr 11	1.1	1.2	1.3	1.1	1.6
Chr 12	1.1	1.2	1.3	1.1	1.6
Chr 18	1.1	1.1	1.2	1.1	1.6
Chr 19	1.3	1.2	1.3	1.3	1.6
Chr X	0.9	1.0	1.0	0.9	1.6

FIGURE 1.13: **Summary of exponents α for different cell lines and experimental approaches** Approximate values of the exponents α of the contact probability across different cell lines (GM06990, GM12878, IMR90, and H1-hESC) and different experiments (Hi-C and TCC; Appendix A) for the whole genome (mean) or specific chromosomes (11, 12, 18, 19, and X). The precise value of α depends on the specific genomic region considered. The exponents reported in this table correspond approximately to the $0.5 - 7.0$ Mb region, as originally discussed in ref. [12] and illustrated in Figs. 1.9 and 1.10.

*Lieberman-Aiden et al. [12].

† Kalhor et al. [16].

‡ Dixon et al. [37].

Chapter 2

Mean-Field Theory of the Symmetry Breaking Model for X Chromosome Inactivation

2.1 The Symmetry Breaking model of XCI

The Symmetry Breaking (SB) model [24, 25] (see 2.1) poses that the BF is an aggregate of regulatory molecules and, on physics grounds, explains why only one is formed. The model predicts the existence of diffusing molecules which can bind each other and the Xic on the Xs. With no need for further assumptions, Statistical Mechanics calculations show that if molecule reciprocal binding energy or concentration are above a critical threshold, molecule binding cooperativity rather than inducing equal saturation of all target regions induces aggregation of a single major cluster on just one X, leaving the other X naked (see below). Such a configuration corresponds to the thermodynamically most likely state for the system. The phenomenon originates from a phase transition, a switch-like thermodynamic process strongly robust to fluctuations. The emerging single aggregate is interpreted as a Blocking Factor (BF) and designates the future active X. In males, the only X has no competitors to bind the BF, and is protected from inactivation by default.

2.2 The Symmetry Breaking mechanisms

Here we discuss a mean-field theory version of the SB model and illustrate its main features by considering a minimal physical system of molecules diffusing in presence of

two identical binding loci. The only assumption is that molecules, having a volume fraction, c, can bind, with multiple valency, each other with an affinity E_0 , and the two DNA loci with affinity E_X . Interestingly, the model can be mapped into a variant of the Ising model of Statistical Physics. We use a mean-field expression for the system Free Energy, $F = E - TS$. The system energy, E, and entropy, S, are approximated by expressions which only depend on average quantities (T is the temperature in Kelvin and k_B the Boltzmann constant). Consider the probability, x_1 (resp. x_2), that a molecule is found around chromosome X1 (resp. X2), and x_3 that it is found elsewhere (with $x_1+x_2+x_3 = 1$). In a mean-field approach [53], by a leading order expansion, the system energy can be written as: $E(x_1, x_2, x_3) = -E_0c^2(x_1^2+x_2^2+x_3^2) - E_X n_0 c(x_1+x_2) - F_0 x_3$. The first term derives from molecule-molecule interactions and is proportional to the probability that two molecules are close to each other. The second term is the binding energy of molecules to the Xs, and we named n0 the number of available binding sites. To have the total F we must consider a third term which takes into account free energy difference with background, i.e., with respect to a state where molecules are not bound around chromosomes. Finally, the mean-field expression for the entropy is $S(x_1, x_2, x_3) = -k_B(x_1 ln x_1 + x_2 ln x_2 + x_3 ln x_3)$. The system equilibrium states correspond to the minima of F , i.e., to the values of x_1 , x_2 and x_3 where $F(x_1, x_2, x_3)$ has a minimum: $\partial F/\partial x_1 = \partial F/\partial x_2 = \partial F/\partial x_3 = 0$ (with $x_1 + x_2 + x_3 = 1$). For sake of clarity, we illustrate the system properties in the E_X , F_0 , and then treat E_X , F_0 as small perturbations. simplified case where E_0 In that limit, the mean-field equations have always the $x_1 = x_2 = x_3 = 1/3$ solution, describing a state where molecules are equally dispersed in the system; this is the k_BT , the solution $x_i = 1/3$ is unstable and three high-T regime. If, however, E_0c^2 new minima arise where the symmetry is broken as one of the xi becomes larger than the others (say, $x_1 > x_2$, $x_3 = 1/3$, and permutations): now molecules are no longer evenly distributed in the system and tend to form a high density self-aggregate in one region. The transition line between those two phases in the (E_0, c) plane is $E_0c^2/k_BT = 1$, i.e., the critical concentration is: $c_c(E_0, T) = k_BT/E_0$. In facts, above c_c , the equilibrium x_is rapidly depart from $(x_1, x_2, x_3) = (1/3, 1/3, 1/3)$ and approach the value $(x_1, x_2, x_3) \simeq (1, 0, 0)$, or any of the other two corresponding to index permutations. The presence of a non-zero molecule-chromosome binding energy, $E_X n_0 c > F_0$, makes the free energy of the states $(x_1, x_2, x_3) \simeq (1, 0, 0)$ or $(x_1, x_2, x_3) \simeq (0, 1, 0)$ smaller than the free energy of the $(x_1, x_2, x_3) \simeq (0, 0, 1)$ state. Yet, the symmetry between the two X chromosomes is broken as $x_1 \neq x_2$: molecules bind mainly around one, randomly chosen chromosome. Summarizing: if E_0c^2/k_BT is below a threshold, the system most probable state corresponds to configurations where molecules are roughly homogeneously dispersed in the system and the Xs are symmetrically bound since $x_2 = x_1$; above that threshold, instead, the X-symmetry is spontaneously broken because $x_1 \neq x_2$, and molecules form

a single major aggregate[24, 25]. Monte Carlo simulations have provided relatively accurate indications about threshold energy/concentrations in real nuclei[24, 25]. Molecule molar concentration, ρ, is simply related to our parameter c : $\rho \sim c/d^3 N_A$, where N_A is the Avogadro number, and d_0 the locus linear size which is a couple of orders of magnitude smaller than the nucleus diameter (i.e., $d_0 \sim 10nm$, corresponding to 30 bp, a value comparable to the typical size of Transcription Factor (TF) binding sites). For E0 values of the order of TFs binding energies, i.e., a few units in $k_B T$, computer simulations predict threshold concentrations around $c \sim 0.01 \sim 0.1\%$ [24, 25]. Hence, molar threshold concentrations are expected to be around $\rho \sim 0.1 - 1\mu$mole/litre, a value close to usual nuclear protein concentrations. Importantly, recent experiments appear to have discovered some of the molecular components of the BF predicted by the SB model, e.g., two Zn-finger proteins, Yy1 and CTCF: they were shown to form a complex, binding selectively one X, and regulating the XCI switch[54]. Their binding regions within the Xic were also shown to be multiple clustered sites, as suggested by the theory (see below).

2.3 Effects of X chromosome deletions

The SB model can be used to predict the effects of DNA deletions. For simplicity of illustration, we assume to be deep in the high-c phase where the system state corresponds to $(x_1, x_2, x_3) \simeq (1, 0, 0)$ or $(0, 1, 0)$ or $(0, 0, 1)$. Since we aim to describe the Blocking Factor (BF) binding to the chromosomes, we have to switch on EX and F_0 , but we treat them as small perturbations to E_0 . In general, the relative probability, P_r , to find the system in state r, where $r = 1, 2, 3$ corresponds to the three states above, is given by the ratio:

$$P_r = \frac{e^{-\beta F_r}}{\sum_{k=1}^{3} e^{-\beta F_r}} \tag{2.1}$$

where F_r is the Free Energy of state r. Suppose now to delete equally on the two Xs (*homozygous deletion*) some of the DNA binding sites for the BF, so that their original (*wild type*) number, n_0, is reduced to a fraction, $f_0 : n_0 \to f_0 n_0$(see 2.2). In each single cell, the BF overall binding energy to chromosome $i = 1, 2$ is correspondingly reduced: $E_i \simeq -cE_X n_0 f_0 e_B f_0$, where $e_B = cE_X n_0 \geq 0$ is the maximal allowed binding energy. This affects the probability, P_i , of the BF to bind X chromosome $i = 1, 2$, i.e., the probability to be in state $r = i = 1, 2$:

$$P_i = \frac{e^{-\beta F_i}}{e^{-\beta F_0} + e^{-\beta E_1} + e^{-\beta E_2}} \tag{2.2}$$

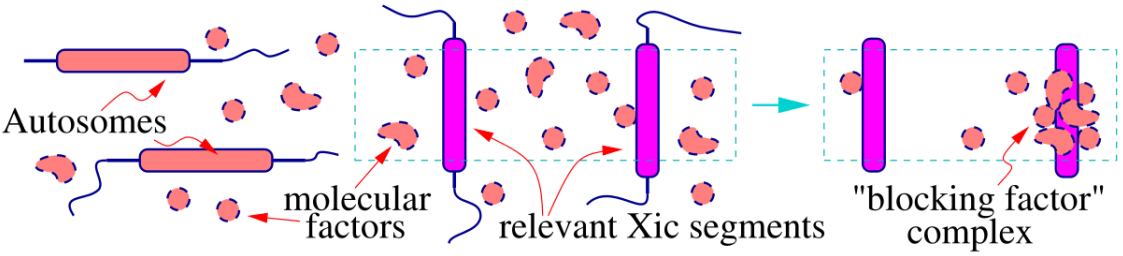

FIGURE 2.1: Pictorial view of the Symmetry Breaking (SB) Model for XCI: a density of molecular factors (produced on sex and non-sex chromosomes, i.e., autosomes) which have a reciprocal affinity, diffuse to interact with two specific Xic segments. If molecule concentration or reciprocal affinity is larger than an appropriate threshold value (of the order of hydrogen bonds) they are shown to spontaneously assemble, as a consequence of a thermodynamic phase transition, into a single major complex attached to only one of the chromosomes. The complex is the 'Blocking Factor' which silences the Xist gene of the bound chromosome.

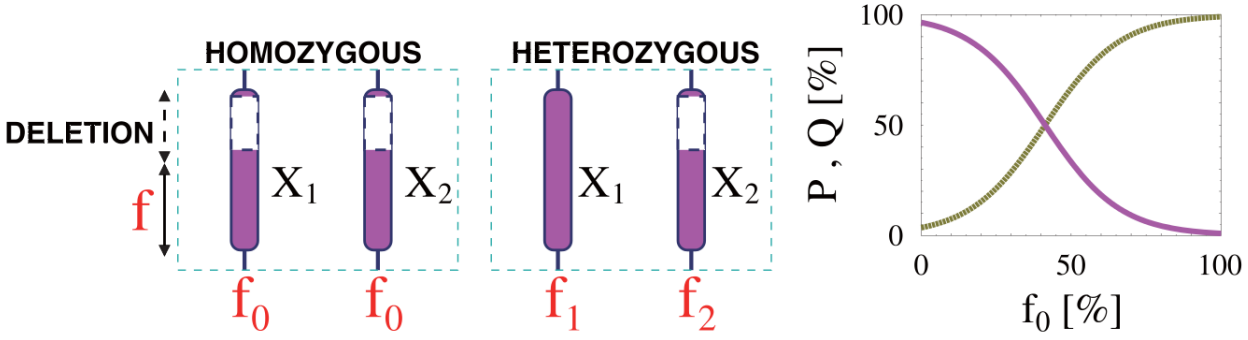

FIGURE 2.2: (color online) Left: picture of X deletions and notation. Right: average probability, $P = P_1 + P_2$, of the Blocking Factor (BF) to bind any of the X chromosomes (purple continuous line) and the probability not to bind, Q (green dashed line), as a function of the fraction, f_0, of DNA binding sites left after a homozygous deletion. Here $\beta F0 = -4$ and $\beta e_B = 8$.

where F_0 is the reference background free energy. So for instance, P_1 is given by the Fermi function:

$$P_1 = \frac{1}{2 + e^{[-\beta(e_B f_0 + F_0)]}} \tag{2.3}$$

and $P2$ is obtained by swapping index 1 with 2 in the above expression.

Consider first the case of long deletions, i.e., when the fraction of remaining binding sites tends to zero, $f_0 \to 0$: in this case P_i falls drastically to its (small) background value, and the BF misses its X targets

$$P_1(f_0 = 0) = P_2(f_0 = 0) = \frac{1}{2 + exp(-\beta F_0)}. \tag{2.4}$$

Such an expression shows that the parameter βF_0 sets the 'background' binding probability of the BF to an X. Note that the requirement $P_1(f_0 = 0) \ll 1$ implies $\beta F_0 \ll 0$. We can also compute the probability, $Q(f_0)$, that in a given cell the BF fails

to bind any of its targets as a function of the deletion extent f_0 :

$$Q(f_0) = \frac{e^{-\beta F_0}}{e^{-\beta F_0} + e^{-\beta E_1} + e^{-\beta E_2}} \simeq \frac{1}{2 + 2e^{[\beta(e_B f_0 + F_0)]}} \tag{2.5}$$

and

$$P_1 = P_2 = \frac{1 - Q}{2}. \tag{2.6}$$

$Q(f_0)$ has its maximum $Q(0) = 1/[1 + 2\exp(\beta F_0)]$ for $f_0 = 0$ and decreases *á* la Fermi with $\beta e_B f_0$ (see 2.2). The opposite behaviour is found for $P_1(f_0)$. Note that the condition $\beta e_B \gg \beta|F_0|$ must be verified for the BF to have a sensible binding probability. For typical values of Transcription Factors binding energies, which are in the weak biochemical scale, say $E_X \sim 5 - 15 k_B T$, such a condition can be more easily verified by increasing the number of the binding sites n_0 . This could explain the presence of clusters of binding sites observed experimentally for, e.g., CTCF and Yy1, which would provide robustness to the system. In a different perspective, $Q(f_0)$ predicts how many cells in a colony die for XCI failure (because the BF is not correctly bound) after a $(1 - f_0)n_0$ long deletion is made in the population DNA. Interestingly, the predicted behaviour is consistent with the available results from deletion experiments[20–22].
A similar calculation can be made when the deletion is not equal on the two Xs (heterozygous deletion). Name f_1 and f_2 the fraction of binding sites left respectively on X_1 and X_2 . Then

$$\frac{P_1}{P_2} = \exp \beta e_B (f_1 - f_2). \tag{2.7}$$

Importantly, such a ratio is different from one if $f_1 \neq f_2$: the BF is preferentially attached to the X with the shorter deletion and $P_1 \neq P_2$ (the skewing being approximately exponential in $f_1 - f_2$), a behaviour observed in heterozygous deletions[20–22]. Finally, this analysis also explains that XCI is affected by thermal effects, i.e., by temperature changes in the experimental conditions.

Chapter 3

Conformation Regulation of the X Chromosome Inactivation Center: A Model

3.1 Introdution

X-Chromosome Inactivation (XCI) is the vital process occurring in female mammalian cells whereby one randomly selected X is transcriptionally silenced to balance dosage with respect to males [20–22, 28]. XCI is regulated by a region on the X chromosome, the X inactivation center (Xic), which encompasses a key group of neighboring non-coding genes (see Fig. 3.1.A) including, e.g., Jpx, Xist, Tsix and Xite [20–22, 28]. The fate of the X is determined by its Xist gene which is strongly upregulated on the future inactive X and repressed on the other X. In turn, Xist is negatively regulated by Xite=Tsix, and positively regulated by Jpx, Rnf [55], and other factors [29, 56, 57]. Before random XCI starts, a complex epigenetic program, coupling transcription and chromatin remodelling [58, 59] to pluripotency factors [55, 60, 61], produces a state where the Xic has the same spatial conformation on the two X chromosomes [55] and both Xist alleles are just weakly active. Upon XCI, an unknown symmetry breaking mechanism determines the opposite behaviour of the two Xist, and induces alternative modifications of the three-dimensional conformation of their Xic [27, 62]. Finally, on the designated inactive X further chromatin reorganizations occur as a heterochromatic compartment forms into which genes are recruited to be silenced [28, 63]. Several molecular factors are known to be involved in the process [22, 28], including noncoding transcripts, chromatin modifiers and organizers, such as CTCF (a Zn finger having arrays of binding sites on the Xic), Dnmt3a, Oct4 and other pluripotency factors [54, 55, 59–61, 64]. Different models have

been proposed to describe random XCI [24, 30, 65–67], but still none to elucidate its associated chromatin changes, whose nature remains mysterious.

3.2 Model

We represent the relevant region of each X chromosome (see scheme in Fig. 3.1.B) by a standard model of polymer physics, a selfavoiding bead chain [68]. In the light of Xic current 3C data [27], we pose that along each polymer there are, for simplicity, two type-α regions which have an array of binding sites for type-A Brownian molecular factors. Each polymer has also two type-β regions with binding sites for a different kind of molecular factors (type-B). Finally, the polymers have a type-γ region whose binding sites can be bound by either type-A or B molecules. Thus, type-A molecules (resp. type-B) can bridge a type-α (resp. type-β) and a type-γ site. For simplicity, with no loss of generality, we consider the case where the two types of molecules have the same concentration, c, and the same affinity, E_X, for all binding regions. Similarly, we assume that type-α and type-β regions have the same number of binding sites, n_0, than type-γ. The value of n_0 is fixed to have a total binding site number of the order of known Xic binding molecules. As CTCF is a general chromatin organizer which has been associated to XCI and its Xic binding sites have been well characterized [54], we use it as an example (and set $n_0 \sim 20$). For simplicity, n_0 is here also the length of the intervening inert sequences between them. Type-A (resp. type-B) molecules can bind, with multiple valency, each other with affinity E_{AA} (resp. E_{BB}); we set $E_{AA} = E_{BB} \equiv E_0$ and, considering the number of binding domains of CTCF, the valency to four. We investigate by Monte Carlo (MC) simulations the conformations of the system as they spontaneously emerge when the three control parameters, (c, E_X, E_0), are varied (Appendix B for details on simulations). For computational purposes, the system lives in a cubic lattice with a lattice spacing d_0, whose value corresponds to the typical size of a DNA binding site, and can be roughly estimated to be $d_0 \sim 10n$m. The volume concentration of molecules in our model, c, can be related to molar concentrations $\rho : \rho \sim c/d_0^3 N_A$, N_A being the Avogadro number (details in Text S1). Thus, for instance, a typical nuclear protein concentration of $\rho \sim 0.1\mu$mole/litre would correspond to $c \sim 10^{-2}\%$. Below we consider concentrations in the range $c \sim 10^{-4} - 10^{0}\%$ and binding energies in the weak biochemical scale (a few units in k_BT). Finally, conversion of MC time unit to real time is obtained by imposing that the diffusion constant of our polymers is of the order of measured chromatin diffusion constants (see Text S1 for details).

3.3 Results

3.3.1 Establishing stable interactions

We first show that diffusing molecules can produce a looped conformation on each polymer where type-α and type-β stably interact with type-γ region. The process is based on a thermodynamic mechanism (a phase transition, in the thermodynamic limit) which acts switch-like when concentration/affinity of binding molecules rise above a threshold [19]. Before describing our MC results in details, we illustrate the underlying mechanisms. A single, say, type A molecule forms a bridge between type-α and type-γ regions via the stochastic double encounter of the molecule with its binding sites. This is, though, an unlikely event, especially if molecule concentration, c (or E_X , see below), is small. And the half-life of such a bridge is short when weak biochemical interactions are considered. Thus, on average the regions float away from each other (see pictorial representation in the bottom panel of Fig. 3.2, "Open State"). At higher c (or E_X), however, many a molecule can bind type-α/c regions and stabilize the conformation via a positive feedback mechanism as their bridges reinforce each other and facilitate the formation of additional bridges. The concentration where such a positive feedback mechanism starts winning marks the threshold above which stable contacts are established (pictorial representation in the bottom panel of Fig. 3.2, "Stable Interaction"). This pictorial scenario summarizes our MC results. For sake of simplicity, we consider first the case where molecule mutual interaction is turned off, $E_0 \sim 0$, and set as initial configuration of the polymers a randomly open conformation. We measure the interaction order parameter, $\pi = (p_A + p_B)/2$, where p_A (resp. p_B) is the probability to have, on a polymer, a contact of a type-α (resp. type-β) with type-γ region. If neither type-α nor type-β regions are in contact with c, the order parameter is zero, $p \sim 0$; if only one pair is stably interacting then $\pi = 1/2$; finally, $\pi = 1$ if both type-α and type-β loops are established. Fig. 3.2 top panel shows the MC time evolution of $\pi(t)$ for two values of c: if c is small, π remains indefinitely close to zero, $\pi = 0$, as no stable contact is statistically possible; instead, if c is high enough, π grows to a value close to one, $\pi \simeq 1$, showing that both the type-α and b loops are formed.

3.3.2 Conformation switch and sharp regulation

In the space of the control parameters, (c, E_X), a sharp line separates the two regimes, as shown in Fig. 3.2 bottom panel: when c or E_X are small, contacts cannot be stable and $p = 0$; conversely, above the transition line the two loops conformation is reliably established on each polymer, and $p \sim 1$. Such a line marks the boundary between

two thermodynamic phases [19]: it corresponds to the point where the entropy loss due to loop formation is compensated by the energy gain obtained from the establishment of the corresponding bridges. The discovery of such a switch-like behaviour can also explain how loop 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 $n_0 E_X$. Thus, non-linear threshold effects in genetic deletion/insertions of the locus exist.

3.3.3 Threshold values in real nuclei

From Monte Carlo results we can predict concentration (or energy) thresholds in real nuclei. For instance, in vitro measures of CTCF DNA binding energies give $E_X \sim 20k_T$, a typical value for TFs [31, 69]: an extrapolation from Fig. 3.2 then predicts a threshold $c_{tr} \sim 10^-3\%$, corresponding to a typical nuclear protein molar concentration $\rho \sim 10^{-2}\mu$mole/litre (see Text S1). Finally, the mechanism leading to stable loop formation has to be fast enough to serve functional purposes. In our model we find that stable interactions are established on scales of the order of minutes (see Fig. 3.2 top panel and Text S1), a range consistent with biological expectations.

3.3.4 Symmetry Breaking mechanism

The mechanism to induce conformational changes illustrated above acts "symmetrically" on the two polymers. Now we show that molecule homotypic interaction, E_0 , can break the polymer symmetry via a different thermodynamic mechanism. More precisely, if E_0 (and c, see below) is above a critical threshold, a single major aggregate of type A molecules and a single one of type B are formed because of homotypic binding cooperativity: in facts, the energy gain in forming a single cluster of A/B molecules (which maximizes the number of possible chemical bonds) compensates, if E_0 is large enough, the corresponding entropy reduction. The single, say, type A aggregate will then randomly bind just one polymer, leaving the other one "naked" (pictorial representation in the bottom panel of Fig. 3.3, "Symmetry Breaking"). Type-A and B aggregates bind opposite polymers because A and B molecules compete for binding sites in the type-γ region. Hence, if a fluctuation increases the presence of, say, A molecules on one polymer, cooperativity tends to favor their assembling at that site and B molecules are expelled; in turn, the depletion of A around the other polymer favors the assembling of

B molecules on it. On the polymer where the A cluster binds the type-γ region, the B-related loci can no longer be stably linked, and their loop opens; the opposite situation happens on the other polymer. The above scenario results from our MC simulations (Appendix B for details on simulations). We measured the symmetry breaking order parameter, $m_A \sim |\rho_A^{(1)} - \rho_A^{(2)}|/(\rho_A^{(1)} + \rho_A^{(2)})$, where $\rho^(i)$ is the average local concentration of A molecules around the type-γ region of polymer $i = 1, 2$. The m_A parameter is close to zero if an equal amount of A molecules is present around the two polymers, whereas it approaches one if the symmetry is spontaneously broken (m_B and $m \equiv (m_A + m_B)/2$ behave analogously). Fig. 3.3 top panel shows the time evolution of $m_A(t)$ from an initial configuration corresponding to the symmetric state (schematic picture in the bottom panel of Fig. 3.3, "Stable Interaction") where each polymer has two stable loops as seen before: if E_0 is small, m_A remains close to zero at all times and the system remains in a symmetric state; 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 (schematic picture in the bottom panel of Fig. 3.3, "Symmetry Breaking"). The phase diagram of Fig. 3.3 bottom panel shows that the symmetry breaking mechanism is switch-like too: in the (c, E_0) space, as soon as a narrow transition line is crossed the system switches from a symmetrical polymer state to a broken polymer symmetry state. More details are in the Text S1. For sake of simplicity, we considered the case where the concentration/DNA affinity of molecules A and B are the same. However, such an assumption does not affect our general results. The only condition for the Symmetry Breaking and Configurational Switch mechanisms to be triggered is that concentration/ interaction energy of both types of molecules rise above the appropriate threshold.

3.3.5 Symmetry Breaking in real nuclei

As far as XCI is concerned, the predicted single B molecule aggregate is interpreted as an Xist repressing factor (a Blocking Factor, BF) and designates the future active X. The A aggregate marks the X where Xist transcription is enhanced and is interpreted as an activating factor (AF). Importantly, the thresholds predicted by our theory for the symmetry breaking mechanisms also fall in the correct biochemical range (see above and Fig. 3.3 bottom panel). The time scale required to break the symmetry in a real nucleus can depend on a number of details. Our MC provides, thus, only a very rough order of magnitude estimate. As shown in Fig. 3.3 top panel, such a time scale is predicted to be around 10 hours, a value of the order of the time required for XCI initiation. In males 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 type-γ region

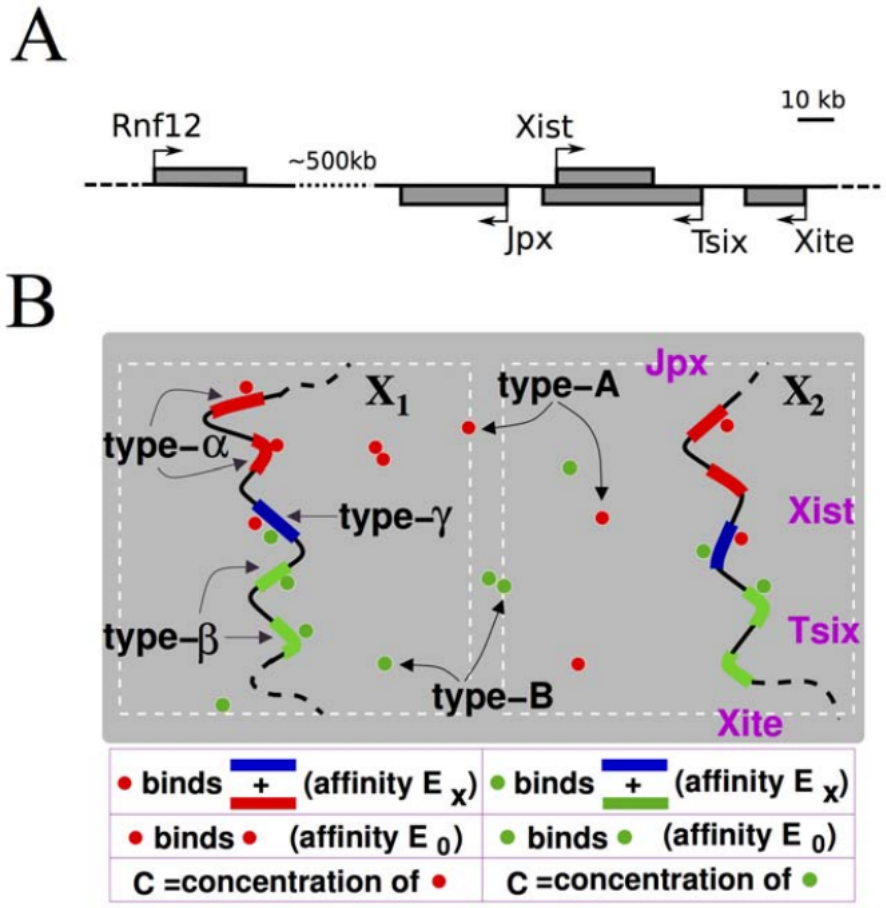

FIGURE 3.1: **The model. Panel A** is an illustration of the region of the X Inactivation Centre (Xic) around the Xist gene. The scheme in **panel B** zooms on the key regions of the two polymer model investigated here. Each polymer has two type-α (red), two type-β (green) and a type-γ (blue) regions. Type-α and type-γ can be bridged by type-A molecules (red circles) with an affinity E_X ; type-β and type-γ by type-B molecules (green circles). Each molecular species has a concentration c. Type-A molecules have also a homotypic mutual interaction of affinity E_0 , and similarly type-B ones. The presumptive mapping areas on the Xic are also illustrated (right panel).

are expected, in general, to be different: $E_{X_A} \neq E_{X_B}$. Hence, if E_{X_B} is larger than E_{X_A} , it is thermodynamically convenient that B molecules bind the X, a difference of a few units in k_T being sufficient to skew of orders of magnitudes the binding probability of A and B. Finally, variants of the model can be considered to account for further biological details. For instance, additional molecular factors, or the effects on polymer colocalization can be discussed (see Text S1), but no relevant changes to the present scenario are found.

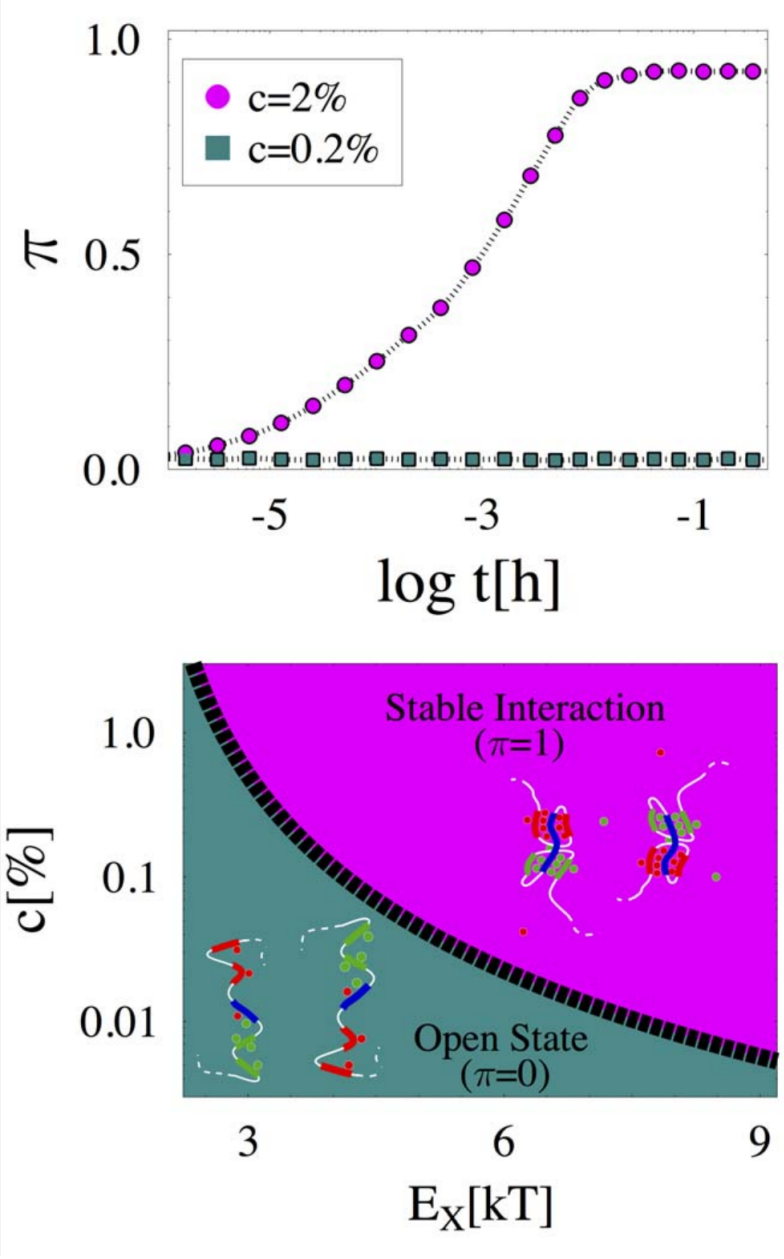

FIGURE 3.2: **Conformational Switches and the establishment of stable interactions. Top Panel** π is the probability of type-α and type-β regions to loop stably onto type-γ region. Its time evolution, from an initial open polymer conformation (see schematic representation in the Bottom Panel, "Open State"), is shown for two characteristic values of the concentration, c (here $E_X \sim 3k_T$ and $E_0 \sim 0$). For $c = 0.2\%$, p is zero at all times: neither type-α nor type-β regions succeed in forming stable contacts with type-γ, and the polymer conformation remains open. For $c = 2\%$, after a transient of the orders of minutes, π approaches one: a stable, looped conformation is established (see schematic picture in the Bottom Panel, "Stable Interaction"). **Bottom Panel** The conformation phase diagram in the (E_X, c) plane is shown (for $E_0 = 0$): in the region below the sharp transition line, c_{tr} (E_X) (black dashed line), the polymers are found in an open state; above c_{tr} (E_X), they exhibit a conformation change, symmetrical on the two polymers, as a stable interaction of type-α and type-β with type-γ region is established.

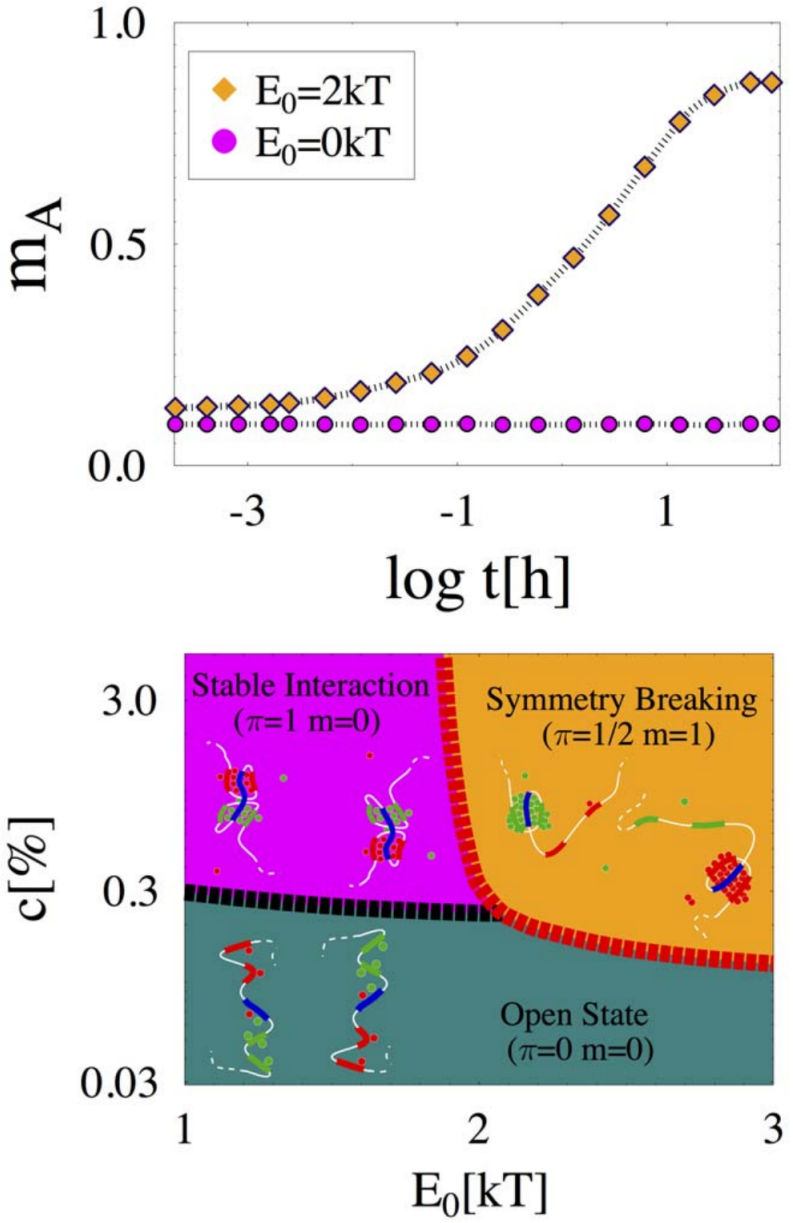

FIGURE 3.3: **The Symmetry Breaking (SB) mechanism. Top Panel** The SB parameter, $m_A \sim |\rho_A^{(1)} - \rho_A^{(2)}|/(\rho_A^{(1)} + \rho_A^{(2)})$, is the normalized average difference of type-A molecule density around type-γ region of polymers 1 and 2. Its dynamics, from the initial symmetrical polymer looped state (as in the schematic picture in the Bottom Panel, "Stable Interaction"), is shown for two characteristic values of molecule homotypic interaction energy, E_0 (here $c = 2\%$ and $E_X = 3k_T$). If $E_0 = 0$, m_A is close to zero: molecules are equally distributed around the polymers. If $E_0 = 2k_T$, after a transient of about ten hours, m_A approaches one, i.e., either $\rho_A^{(1)} \to 0$ or $\rho_A^{(2)} \to 0$: molecules have aggregated around only one of the polymers, and their binding symmetry is broken (as in the schematic picture in the Bottom Panel, "Symmetry Breaking"). **Bottom Panel** The phase diagram in the (E_0, c) plane (for $E_X = 3k_T$) has three phases. If E_0 is below the transition line, $E_{tr}^{SB}(c)$ (red dashed line), the system is in one of its symmetric phases: the "Open State" phase (at low c) or the symmetrical "Stable Interaction" phase. If $E_0 > E_{tr}^{SB}(c)$, the conformational symmetry isbroken ("Symmetry Breaking" phase): the type-α loop persists only on one randomly chosen polymer, and type-β on the other.

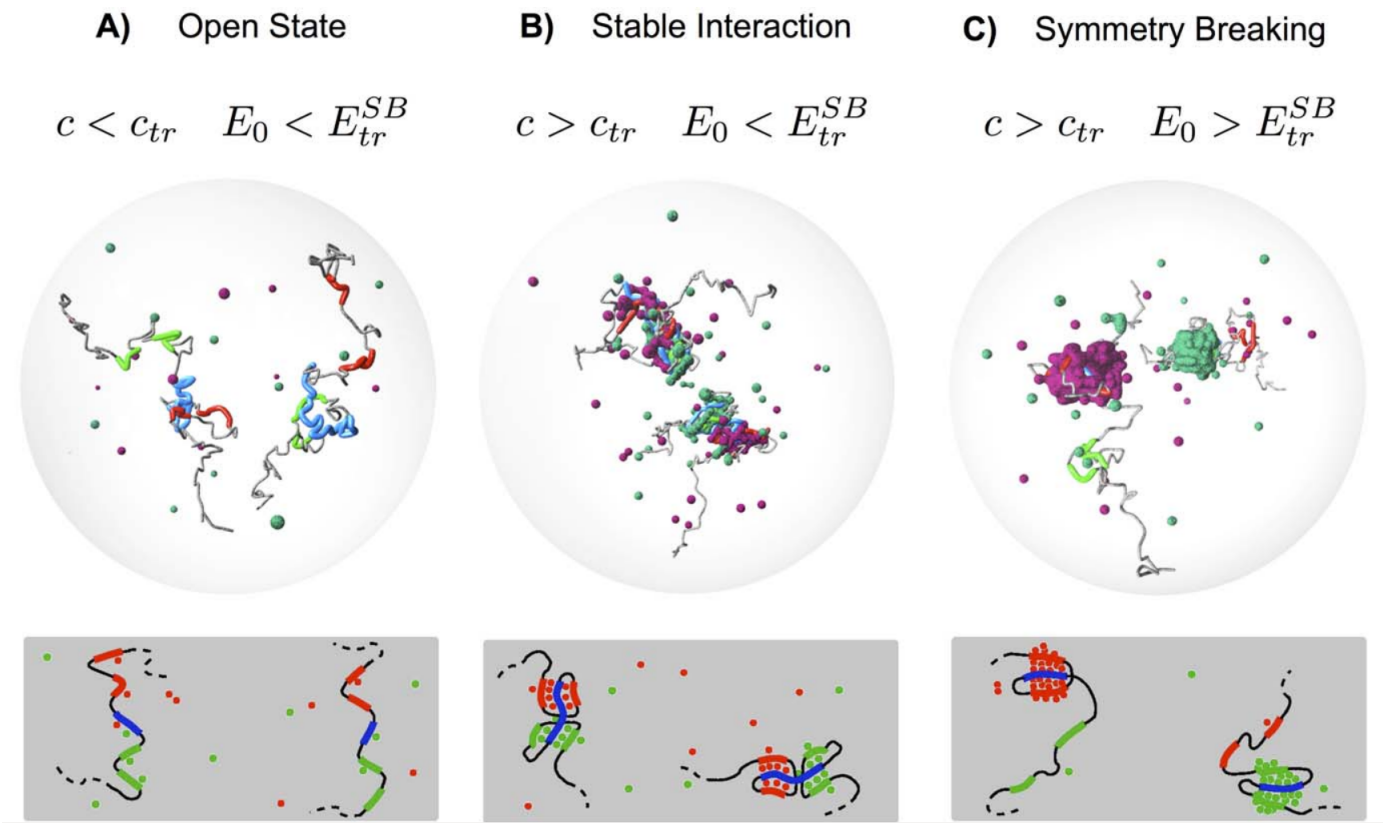

FIGURE 3.4: **System states and transitions**. The figure summarizes the system possible states and how they change by action of the Conformation and the Symmetry Breaking switch (top pictures are from MC simulations, bottom ones are schematic drawings). The switches have a thermodynamic nature and are regulated by increasing, e.g., c and E_0 (i.e., molecule concentration and homotypic interaction) above precise threshold values, c_{tr} and E_{tr}^{SB}. **A)** For cvctr and $E_0 < E_{tr}^{SB}$, the polymers are found in a random open state. **B)** For $c > c_{tr}$, a conformation change is activated: type-α and type-β regions stably interact with type-γ, and a two loop conformation is established symmetrically on the two polymers. **C)** If $E_0 > E_{tr}^{SB}$, a symmetry breaking occurs as the type-α loop persists on one, randomly selected, polymer (where type-β loop is released), whereas the other polymer takes the opposite conformation. This results from the self-assembling of a single major aggregate of type-A and of type-B molecules competing to bind to type-γ region.

3.4 Discussion

We now discuss how the present scenario can recapitulate in a unified framework important experimental results on XCI.

3.4.1 Xic architecture, "counting" and "choice"

Before XCI, the Xic conformation is found to be identical on the two X's [27]: Tsix and Xite genes are looped onto a "buffer" region; similarly, Jpx, Xist and the "buffer" form a second hub with Xist. Upon XCI, on the future active X, the Jpx-Xist-buffer hub

opens while Xite remains in contact with Tsix. On the other X, instead, the Tsix-Xite interactions is lost whereas Xist and Jpx remain in contact. Our model rationalizes how those elements are sharply regulated to recognize each other and to form stable interactions based on weak biochemical bonds. It can also explain how the same physical elements later at XCI spontaneously break the X symmetry. The molecular aggregate bound, in our model, to the type-β regions (which should encompass the Tsix-Xite area) is interpreted as a factor related to Xist silencing (i.e., to its Blocking Factor, BF [20–22]) and designates the future active X; the different aggregate bound to type-α regions, encompassing the Jpx area of the other X would be linked to an Xist activating factor (AF) [29, 56, 67]. The link between architectural changes and choice of fate emerges here naturally. During XCI establishment, the inactive X undergoes further architectural reorganization [28, 62, 63]. The mechanistic details of those conformational changes are still not understood, but they could involve mechanisms as those illustrated here. Other interesting models have been proposed for "counting&choice" at XCI, but still none had focused on the Xic spatial organization, including our original Symmetry Breaking theory [24]. In the approach of ref. [30], each X chromosome is assumed to have an independent probability to initiate inactivation. Two competing factors exist: an X-linked XCI-activator and an XCI-inhibitor produced by autosomes. In a male XY cell the XCI-activator concentration is too low to initiate the inactivation of the only X; in female XX cells the initial XCI-activator concentration is, instead, above the threshold needed to start XCI. As soon as one X is inactivated, the XCI-activator concentration falls down to the levels found in males, and thus the other X remains active. A different model [67] poses that two types of sites are present on the X: "XCI-init" which is responsible for the initiation of inactivation of the X bearing it, and "XCI-repres" sites which inhibit the action of "XCI-init". Each active X produces molecules, say A molecules, which bind to some autosomal sites. If these sites are saturated, the autosomes produce a set of molecules I, which, with a "Symmetry Breaking" mechanism [24], self-assemble into a single molecular factor and inhibit the activity of "XCI-repres" sites on one of the two X, determining its inactivation. As the availability of the A signal is reduced, it is no longer sufficient to saturate the autosomal receptors, and the remaining X remains active. The mechanisms for conformational changes we discussed here are rooted in thermodynamics and are, thus, very robust to difference in molecular details. They could apply then to all the mentioned models for "counting&choice". An interesting question concerns the applicability of those models to mammals other than mice. Important differences have emerged, for instance, between human and mice XCI [70, 71]. As stated above, the mechanisms we discussed for Xic architecture in mice stem cells are very robust, yet data on other organisms are still too scarce to decide whether such mechanisms might apply elsewhere.

3.4.2 Xic deletions/insertions and XCI

The phenotype of key deletions along the Xic (see reviews in [20–22, 24, 67] and ref.s therein) can be explained by our model. The D65kb deletion [72] removes 65kb encompassing Xite and part of Xist/Tsix. In heterozygous females the deleted X is always inactivated. In males it leads to the inactivation of the only X; the shorter the deletion considered within the D65kb (see DAS, DAJ, DAV, D34 [73]), the smaller the fraction of ectopic X inactivations in a population. Those deletions, in our model, map into sites where the Xist "blocking factor" (BF) binds (and blocks inactivation of that X): D65kb removes a large portion of binding sites, thus the deleted X has a strongly reduced affinity for the BF (w.r.t. the wild type X) which does not bind there; the shorter the deletion, the weaker the effect. So, in heterozygously deleted females a skewed random XCI occurs, whereas in males the only X can be inactivated. These deletions can also impact the formation of the BF itself because the involved regions possibly encode some of its components. Heterozygous TsixDCpG [74] and XiteDL [75] deletions in females also result in the inactivation of the deleted X. Their homozygous counterpart produces, though, "aberrant counting/chaotic choice", i.e., presence of two active or inactive X's in a fraction of the cell population [65]. While that cannot be easily rationalized by other models (see, e.g., [30]), in our framework it is originated simply because the BF can fail to bind at all [25]. DXTX is deletion including Xist, Tsix and Xite, which in heterozygous causes a skewed XCI, as only the Wild Type X gets inactivated [30]. In the frame of our model DXTX could have a double effect: on the one hand, it hinders the binding of the AF and BF to the deleted X, by removing a number of their binding sites; on the other it affects especially the BF, since it removes the Tsix=Xite genes which are presumably linked to some of the BF components. Thus, the overall effect will be that while the deleted X remains active (as it lacks Xist), the BF is depleted and the AF wins the competition for binding the Wild Type chromosome, which is then inactivated. Transgenic insertions are also interesting [76]. One of the predictions of our model is the highly non-linear effect of deletion/ insertion, due to the "switch-like" nature of the underlying thermodynamic mechanism. The insertion experiments of ref. [76] support this view: long Xic transgenes can cause inactivation on male ES cells only when they are present in multiple copies, while single insertions do not have appreciable effects. The outcome of other deletions/insertions, such as XistDpromoter [77], XistD1-5 [78], Jpx [29], Rnf 12 [56], etc., are similarly explained (see Text S1). XCI in diploid cells with more than two X and in polyploid cells [30] can be understood as well in our scenario (see Text S1), but additional biological hypotheses are required, since key pieces of information are still missing. In summary, we illustrated physical switch-like mechanisms establishing conformational changes and symmetry breaking in a polymer model. For clarity, we included just the required minimal ingredients, but our model can

accommodate more realistic molecular details. It can be mapped into the Xic region of X chromosomes to explain their complex self-organization and other important aspects of random XCI, such as the deep connection between Xic architectural changes and Xist choice of fate, reconciling within a single framework a variety of experimental evidences. The on-off character of the underlying mechanisms can also explain how sharp and reliable regulation of XCI can be attained by simple strategies, such as gene upregulation or chromatin modification. It supports a picture where random XCI could be governed by a few core molecular elements and basic physical processes. Two main groups of molecular factors are envisaged to control the process and to produce an activating and a blocking factor for Xist. The specific polymer regions in our model emerge as key cis- regulators which orchestrates functional contacts along the Xic. Experiments targeted at that area could test their role. The model also predicts threshold effects of, e.g., genetic deletions of the regulatory regions. The precise nature of factors and sequences involved at XCI could differ from the minimal one considered here, yet the thermodynamic mechanisms we discussed are robust and independent of the specific molecular details. Similar mechanisms could be, thus, relevant to XCI and, more generally, to other nuclear processes requiring, for example, chromatin spatial reorganizations [1, 2, 79] or alternative choices [80]. Xic deletions/insertions and XCI The phenotype of key deletions along the Xic (see reviews in [20–22, 24, 67] and ref.s therein) can be explained by our model. The D65kb deletion [72] removes $65kb$ encompassing Xite and part of Xist/Tsix. In heterozygous females the deleted X is always inactivated. In males it leads to the inactivation of the only X; the shorter the deletion considered within the D65kb (see DAS, DAJ, DAV, D34 [73]), the smaller the fraction of ectopic X inactivations in a population. Those deletions, in our model, map into sites where the Xist "blocking factor" (BF) binds (and blocks inactivation of that X): D65kb removes a large portion of binding sites, thus the deleted X has a strongly reduced affinity for the BF (w.r.t. the wild type X) which does not bind there; the shorter the deletion, the weaker the effect. So, in heterozygously deleted females a skewed random XCI occurs, whereas in males the only X can be inactivated. These deletions can also impact the formation of the BF itself because the involved regions possibly encode some of its components. Heterozygous TsixDCpG [74] and XiteDL [75] deletions in females also result in the inactivation of the deleted X. Their homozygous counterpart produces, though, "aberrant counting/ chaotic choice", i.e., presence of two active or inactive X's in a fraction of the cell population [65]. While that cannot be easily rationalized by other models (see, e.g., [30]), in our framework it is originated simply because the BF can fail to bind at all [25]. DXTX is deletion including Xist, Tsix and Xite, which in heterozygous causes a skewed XCI, as only the Wild Type X gets inactivated [30]. In the frame of our model DXTX could have a double effect: on the one hand, it hinders the binding of the AF and BF to the deleted X, by removing a number of their binding sites; on the other it affects especially the BF, since it removes

the Tsix=Xite genes which are presumably linked to some of the BF components. Thus, the overall effect will be that while the deleted X remains active (as it lacks Xist), the BF is depleted and the AF wins the competition for binding the Wild Type chromosome, which is then inactivated. Transgenic insertions are also interesting [76]. One of the predictions of our model is the highly non-linear effect of deletion/ insertion, due to the "switch-like" nature of the underlying thermodynamic mechanism. The insertion experiments of ref. [76] support this view: long Xic transgenes can cause inactivation on male ES cells only when they are present in multiple copies,

Discussion

Interphase nuclei exhibit dynamic chromatin structures that change in response to cellular signals and influence patterns of gene expression. We show that the SBS model can capture the key aspects of chromatin folding behaviors detected experimentally across different cell systems and by different technical approaches. The model describes how genomic architectures can spontaneously arise with a switch-like nature that can explain how a sharp regulation of nuclear architecture can be obtained reliably by simple strategies, such as protein up-regulation or modification, without the need to fine tune these specific parameters. Under different initial conditions, the polymer displays a variety of transient conformations that evolve into specific stable states (Fig. 1.5): open polymers, closed polymers, and intermediate fractal states. In our scenario, the open polymer state represents open euchromatin ($\nu \sim 0.58$, $\alpha \sim 2.1$), whereas the compact state describes dense heterochromatin ($\nu \sim 0$, $\alpha \sim 0$). The region around the threshold fractal state includes states with exponents that fit with Hi-C-averaged data from human cells (α of approximately $0.9 - 1.6$; Table S1), but also from Drosophila embryos (α of approximately 0.70 or 0.85 for open and closed chromatin, respectively) [13]. However, Hi-C (see Appendix A) data inherently represent average behaviors across a population of cells and chromosomal loci. Although methodological variations could potentially be responsible for differences observed between datasets, the comparison of specific chromosomes within datasets yielded consistent behaviors, such as the deviation of chromosomes 18, 19, and X from the average genome behavior. Thus, our analysis strongly supports the conclusion that the principles of chromatin folding in interphase nuclei cannot be recapitulated by a single "universal" conformational state (and its given α). The simple SBS model considered here illustrates key physical concepts and basic required ingredients to explain chromatin folding in a variety of states identified in living systems. Although specific molecular details can be incorporated into more complex versions of the model (such as the presence of different binders or nonhomogeneous distributions of binding sites), the general range of folding behaviors will remain the same. Many complications arise in real nuclei, including chromatin entanglement effects that are resolved through the action of topoisomerases, self-interactions beyond steric hindrance, and interactions

with the nuclear lamina. In reality, a variety of specific binding factors exist, and thus a complexity of folding states is present inside cell nuclei, where different regions can spontaneously fold into different chromatin states. Importantly, polymer scaling theory ensures that the exponents in $R^2(s)$ and $P_c(s)$ are independent of the minute details of the system considered and reflect universal properties [33]; these parameters are not affected by detailed mapping onto real chromosomes (e.g., the chosen coarse graining level used in the polymer models and the size of binding sites). Therefore, the general structural properties of our model are relevant to real chromatin. It will be interesting in the future to use the SBS model to explore the behavior of two or more chromosomes when they are constrained in the cell nucleus. As additional genome-wide data become available, important issues that can be addressed with the SBS model include the extent and dynamics of chromatin intermingling and the effects of steric hindrance between chromosomes. The SBS model can explain the nature of the mechanisms underlying chromatin self-organization whereby nuclear architecture is governed by a few core molecular ingredients and basic physical processes. More generally, the thermodynamic mechanisms discussed, which are robust and independent of specific molecular details, will be relevant to many cellular and nuclear processes requiring spatial organization [1, 2, 43].

In [23] (Chapter 2) we have considered a 'mean-field' treatment of the Symmetry Breaking (SB) model of the X-Chromosome Inactivation (XCI) phenomenon, the vital process whereby one X, randomly chosen, is silenced to compensate dosage of X products with respect to males. This model is an example of applications[18, 19, 24] of Statistical Mechanics to Molecular Biology. The SB picture explains two major routes to regulate XCI: changes in E_0 or c. This illustrates on quantitative grounds how chromatin modifications, such as DNA methylation or post-translational modifications of DNA binding proteins or changes in molecule concentrations (well described cell strategies), can result in dramatic, switch-like, regulatory effects. It recapitulates the phenotype of a number of genetic deletion/insertion experiments[65, 72, 74, 75, 81–83] and can predict in details "skewing" and "chaotic/aberrant" effects in homozygous and heterozygous deletions of the Xic[24, 25, 84]. The basic model can accommodate further elements, such as both activating and inhibiting factors. And other, redundant, regulatory mechanisms may well occur in XCI. The SB mechanism, though, for its simple and robust physical grounds, could provide switch-like regulation also in other biological processes. The development of a conceptual understanding of XCI could help the rational design of strategies to attack a variety of important and still largely unexplained phenomena involving a form of stochastic regulation, such as Random Monoallelic Expression. XCI is only the best studied case of this class which includes about 10% of our genes[80], 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.

Finally in [26] (Chapter 3) by computer simulations we showed how an SBS-like model (Fig. 3.1.B) predicts that two kinds of molecular regulators, type-A and B molecules, interact with a set of specific regions along the polymers. Current 3C data [27] (Appendix A) suggest that our type-α and type-β regions map respectively in the area 5' and 3' to Xist, while the type-γ region is in between. We showed that in our model only three classes of stable conformational states exist (see Fig. 3.4 A,B,C). The system spontaneously falls in one of them, according to molecule concentration and homotypic interaction, c and E_0. State changes are regulated by a "conformation" and by a "symmetry breaking" switch, related to two distinct thermodynamic phase transitions [85]. The switches are controlled by changing c and E_0 above/below specific threshold values. Their on/off nature can explain once again how a sharp regulation of nuclear architecture and stochastic choice of fate can be reliably 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 μmole/litre concentrations).

Appendix A

Experimental Technique

A.1 Chromosome Conformation Capture (3C) techniques

3C is a molecular method to investigate whether certain genomic regions are in close proximity in three-dimensional nuclear space. Cells are fixed with formaldehyde so that chromatin that is close together is cross-linked. The chromatin is digested with a restriction enzyme and ligated at low DNA concentration to promote intra-molecular ligation between the cross-linked pairs of DNA fragments. After the cross-links are reversed, quantitative PCR is used to interrogate whether two DNA sequences were closely associated in the nucleus, with the number of ligations giving a measure of the frequency of interaction.

3C carbon copy (5C) is a high-throughput adaptation of 3C that uses a multiplex ligation-mediated amplification (LMA) step to generate a library of ligation junctions formed during the 3C procedure. The composition of the 5C library is then determined by large-scale sequencing or by microarray analysis. Thousands of primers can be used in the LMA step, which allows mapping of extensive networks of DNA sequence interactions throughout the genome.

3C-on-ChIP (4C) allows for an unbiased, genome-wide search for DNA loci that closely associate with a given sequence. Ligation products produced by 3C are trimmed by digestion with a frequent cutter and circularized, so that an inverse PCR from the 'bait' locus will amplify all its interacting partners. These partners can then be identified by large-scale sequencing or by microarray analysis. All of these 3C-based methods need careful controls and quantification ([86]).

A.2 Hi-C Technique

Hi-C techique (Fig. A.1) is an upgrading of 3C techniques that is able to produce a genome-wide contact matrix (Fig. A.2) from which we estimate the experimental $P_c(s)$.

A.3 Hi-C and TCC data analyses

Hi-C data [12] from GM06990 cells mapped to hg18 were downloaded from the Gene Expression Omnibus (dataset GSE18199, file GSE18199_RAW.tar). The data include three GM06990 replicates (HindIII, HindIII Biological Repeat, NcoI). The three GM06990 replicates were concatenated to calculate the contact probabilities genome-wide and for individual chromosomes using a similar method to the one described by Lieberman-Aiden et al. [12]. Hi-C and TCC paired-end reads from GM12878 cells [16] were downloaded from the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA, entry SRA025848). The data include two sequencing runs each of HindIII Hi-C and TCC replicates. Hi-C paired-end reads from IMR90 and H1-hESC cells [37] were downloaded from SRA (entry SRP010370). The IMR90 data represent eight sequence runs from two HindIII replicates, and the H1-hESC data are from seven sequencing runs of two replicates. The sequence data for a given cell line were first combined, and the raw paired-end reads were mapped separately to hg18 with Bowtie [87]. Mapped reads were then joined and processed to remove all but one instance of duplicate read pairs. To calculate contact probabilities P_c , intrachromosomal contacts were first extracted and sorted based on chromosome number. Pairwise contacts were then sorted by genomic distance into incremental bins of 100 kb, and the number of Hi-C sequencing reads in each bin was calculated. The contact probability Pc at a given chromosomal distance (s) was determined by dividing the Hi-C read count at that distance (X_d) by its corresponding normalization constant (N_c) as follows: $P_c(s) = X_d/N_c(s)$. N_c corresponds to the total number of possible site pairs separated by a given s along a chromosome as follows: $N_c(s) = L_c s$. N_c values were determined for each chromosome by subtracting L_d from the total chromosome length (L_c) and calculating the sum of all possible pairs separated by that chromosomal distance. Genome-wide $P_c(s)$ values were calculated by dividing the sum of read counts by the sum of expected probabilities from all chromosomes at a given s.

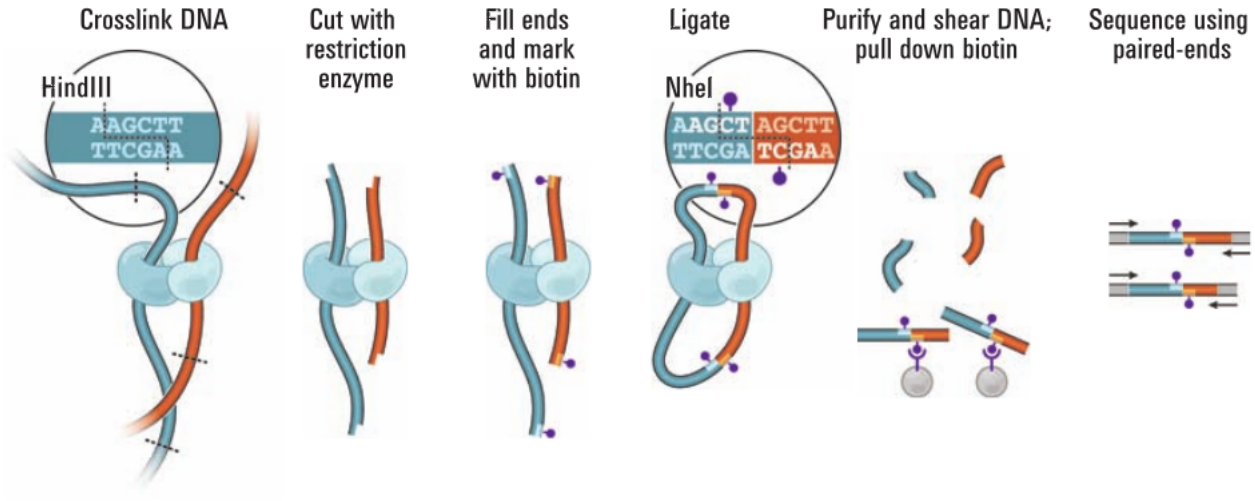

FIGURE A.1: **Overview of Hi-C (fig. 1A and caption from [12]).** Cells are cross-linked with formaldehyde, resulting in covalent links between spatially adjacent chromatin segments (DNA fragments shown in dark blue, red; proteins, which can mediate such interactions, are shown in light blue and cyan). Chromatin is digested with a restriction enzyme (here, HindIII; restriction site marked by dashed line), and the resulting sticky ends are filled in with nucleotides, one of which is biotinylated (purple dot). Ligation is performed under extremely dilute conditions to create chimeric molecules; the HindIII site is lost and an NheI site is created. DNA is purified and sheared. Biotinylated junctions are isolated with streptavidin beads and identified by paired-end sequencing.

A.4 DNA FISH and Cryo-FISH Technique

DNA FISH uses probes that detect specific DNA sequences in nuclei after denaturation to vizualize the spatial relationships between different chromosomes, or different genes or alleles, or a gene and its distant regulatory elements. FISH can also be combined with immunostaining, which allows visualization of the relationship between genomic regions and nuclear proteins or bodies. Multiple image stacks of the nuclei are taken in the z-axis and are then reconstructed in three dimensions.

Cryo-FISH uses ultra-thin $(100 - 200m\text{m})$ cryosections from sucrose-embedded fixed tissues; this eliminates out-of-focus light scatter that comes from above and below the focal plane in traditional non-sectioned samples [86].

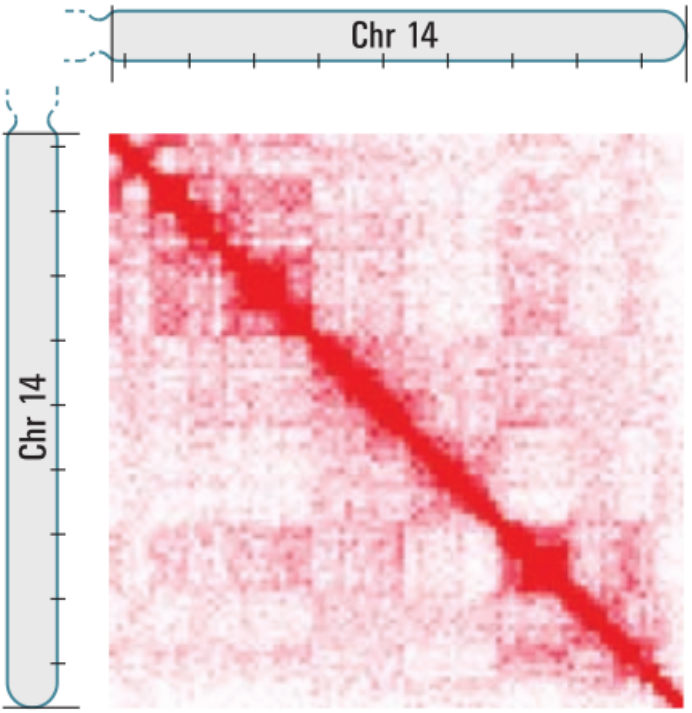

FIGURE A.2: **Hi-C results: Contact Matrix (fig. 1B and caption from [12]).** The submatrix shown here corresponds to intrachromosomal interactions on chromosome 14. Each pixel represents all interactions between a 1-Mb locus and another 1-Mb locus; intensity corresponds to the total number of reads (0 to 50). Tick marks appear every 10 Mb.

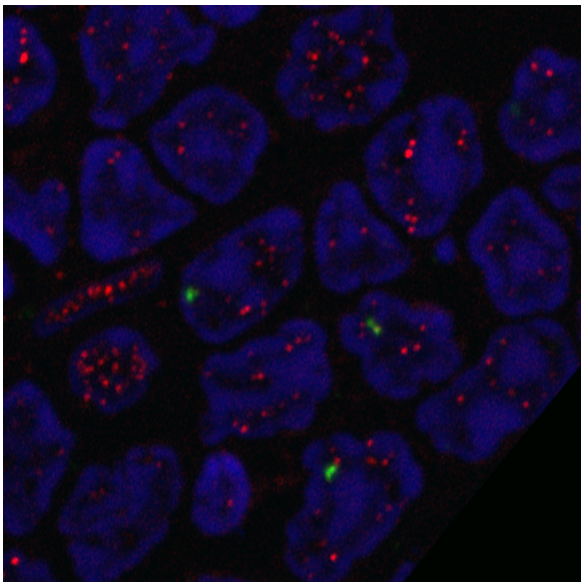

FIGURE A.3: **Cryo-FISH Section.** Red and Green spots are two different types of FISH probes (Picture is courtesy of Ana Pombo's Group at MRC, Imperial College, London).

Appendix B

Simulations Details

B.1 Monte Carlo simulations

In our MC simulations, we employ a Metropolis algorithm and periodic boundary conditions to reduce boundary effects [50]. Each MC step moves every single molecule or polymer bead in random order once, on average, with a transition probability proportional to $\exp(-\Delta H/k_BT)$ [50], where ΔH is the energy barrier encountered in the move. The rate of each trial is given by the Arrhenius factor $r_0 \sim \exp(-\Delta H/k_BT)$, where r_0 is the bare reaction rate. We run our algorithm long enough to approach its stationary state, where the relaxation of the normalized gyration radius as a function of MC time is reported (in the case where $c_m = 25n\text{mol}/L$). The squared radius of gyration is defined as: $R_g^2 = 1/[2n(n-1)] \sum_{i,j=1,n}(\mathbf{r}_i - \mathbf{r}_j)^2$, where r_i is the position of bead $i \in 1, ..., N$. Here, we normalize Rg by the average squared gyration radius of a randomly floating self-avoiding walk (SAW) chain of size n. Pictorially, Rg is the radius of an average sphere enclosing the polymer: It attains a maximum when the polymer is open in a SAW conformation, and a minimum when it becomes a compact globule. We also record the polymer con- tact matrix, a matrix whose entries are defined by the ratio $n_{i;j}/n$, which is the average relative number of times sites i and j along the polymer enter in contact (i.e., are within a distance of $2.5d_0$ from each other). Analogously, the average contact probability, $P_c(s)$, of two sites is defined as having a distance s along the polymer.

Bibliography

[1] C. Lanctot, T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *17230197*, 8(2):104–112, 2007. URL http://hal.archives-ouvertes.fr/hal-00137143.

[2] Tom Misteli et al. Beyond the sequence: cellular organization of genome function. *Cell*, 128(4):787, 2007.

[3] Peter R Cook. The organization of replication and transcription. *Science*, 284 (5421):1790–1795, 1999.

[4] Christopher L Woodcock and Rajarshi P Ghosh. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology*, 2(5), 2010.

[5] Celine Moorman, Ling V Sun, Junbai Wang, Elzo de Wit, Wendy Talhout, Lucas D Ward, Frauke Greil, Xiang-Jun Lu, Kevin P White, Harmen J Bussemaker, et al. Hotspots of transcription factor colocalization in the genome of drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 103(32):12027–12032, 2006.

[6] CH Eskiw, NF Cope, I Clay, S Schoenfelder, T Nagano, and P Fraser. Transcription factories and nuclear organization of the genome. In *Cold Spring Harbor symposia on quantitative biology*. Cold Spring Harbor Laboratory Press, 2011.

[7] Frédéric Bantignies and Giacomo Cavalli. Polycomb group proteins: repression in 3d. *Trends in Genetics*, 27(11):454–464, 2011.

[8] D Peric-Hupkes and B van Steensel. Role of the nuclear lamina in genome organization and gene expression. In *Cold Spring Harbor symposia on quantitative biology*. Cold Spring Harbor Laboratory Press, 2011.

[9] Julio Mateos-Langerak, Manfred Bohn, Wim de Leeuw, Osdilly Giromus, Erik MM Manders, Pernette J Verschure, Mireille HG Indemans, Hinco J Gierman, Dieter W Heermann, Roel van Driel, et al. Spatially confined folding of chromatin in the

interphase nucleus. *Proceedings of the National Academy of Sciences*, 106(10):3812–3817, 2009.

[10] Suchit Jhunjhunwala, Menno C van Zelm, Mandy M Peak, Steve Cutchin, Roy Riblet, Jacques JM van Dongen, Frank G Grosveld, Tobias A Knoch, and Cornelis Murre. The 3d structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell*, 133(2):265–279, 2008.

[11] Christian Münkel, Roland Eils, Steffen Dietzel, Daniele Zink, Carsten Mehring, Gero Wedemann, Thomas Cremer, and Jörg Langowski. Compartmentalization of interphase chromosomes observed in simulation and experiment. *Journal of molecular biology*, 285(3):1053–1065, 1999.

[12] Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950):289–293, 2009.

[13] Tom Sexton, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, Giacomo Cavalli, et al. Three-dimensional folding and functional organization principles of the drosophila genome. *Cell*, 148(3):458, 2012.

[14] T Cremer and C Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews genetics*, 2(4):292–301, 2001.

[15] AIu Grosberg, SK Nechaev, and EI Shakhnovich. The role of topological limitations in the kinetics of homopolymer collapse and self-assembly of biopolymers]. *Biofizika*, 33(2):247, 1988.

[16] Reza Kalhor, Harianto Tjong, Nimanthi Jayathilaka, Frank Alber, and Lin Chen. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature biotechnology*, 2011.

[17] Mariano Barbieri, Mita Chotalia, James Fraser, Liron-Mark Lavitas, Josée Dostie, Ana Pombo, and Mario Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences*, 109(40):16173–16178, 2012.

[18] Mario Nicodemi, Barbara Panning, and Antonella Prisco. A thermodynamic switch for chromosome colocalization. *Genetics*, 179(1):717–721, 2008.

[19] Mario Nicodemi and Antonella Prisco. Thermodynamic pathways to genome spatial organization in the cell nucleus. *Biophysical journal*, 96(6):2168–2177, 2009.

[20] Philip Avner and Edith Heard. X-chromosome inactivation: counting, choice and initiation. *Nature Reviews Genetics*, 2(1):59–67, 2001.

[21] Anton Wutz and Joost Gribnau. X inactivation xplained. *Current opinion in genetics & development*, 17(5):387–393, 2007.

[22] Jeannie T Lee. Lessons from x-chromosome inactivation: long ncrna as guides and tethers to the epigenome. *Genes & development*, 23(16):1831–1842, 2009.

[23] Antonio Scialdone, Mariano Barbieri, Deborah Pallotti, and Mario Nicodemi. Mean-field theory of the symmetry breaking model for x chromosome inactivation. *Progress of Theoretical Physics Supplement*, 191(Supplement 1):40–45, 2011.

[24] Mario Nicodemi and Antonella Prisco. Symmetry-breaking model for x-chromosome inactivation. *Physical review letters*, 98(10):108104, 2007.

[25] Mario Nicodemi and Antonella Prisco. Self-assembly and dna binding of the blocking factor in x chromosome inactivation. *PLoS computational biology*, 3(11):e210, 2007.

[26] Antonio Scialdone, Ilaria Cataudella, Mariano Barbieri, Antonella Prisco, and Mario Nicodemi. Conformation regulation of the x chromosome inactivation center: A model. *PLoS computational biology*, 7(10):e1002229, 2011.

[27] Chia-Lun Tsai, Rebecca K Rowntree, Dena E Cohen, and Jeannie T Lee. Higher order chromatin structure at the x-inactivation center via looping dna. *Developmental biology*, 319(2):416–425, 2008.

[28] Jennifer Chow and Edith Heard. X inactivation and the complexities of silencing a sex chromosome. *Current opinion in cell biology*, 21(3):359–366, 2009.

[29] Di Tian, Sha Sun, and Jeannie T Lee. The long noncoding rna, jpx, is a molecular switch for x chromosome inactivation. *Cell*, 143(3):390–403, 2010.

[30] Kim Monkhorst, Iris Jonkers, Eveline Rentmeester, Frank Grosveld, and Joost Gribnau. X inactivation counting and choice is a stochastic process: evidence for involvement of an x-linked activator. *Cell*, 132(3):410–421, 2008.

[31] Mario Renda, Ilaria Baglivo, Bonnie Burgess-Beusse, Sabrina Esposito, Roberto Fattorusso, Gary Felsenfeld, and Paolo 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. *Journal of Biological Chemistry*, 282(46):33336–33345, 2007.

[32] Ana Pombo, Dean A Jackson, Michael Hollinshead, Zhengxin Wang, Robert G Roeder, and Peter R Cook. Regional specialization in human nuclei: visualization of discrete sites of transcription by rna polymerase iii. *The EMBO journal*, 18(8): 2241–2253, 1999.

[33] Pierre Gilles De Gennes. *Scaling concepts in polymer physics*. Cornell university press, 1979.

[34] Ronald Hancock. Packing of the polynucleosome chain in interphase chromosomes: evidence for a contribution of crowding and entropic forces. In *Seminars in cell & developmental biology*, volume 18, pages 668–675. Elsevier, 2007.

[35] Lindsay S Shopland, Christopher R Lynch, Kevin A Peterson, Kathleen Thornton, Nick Kepper, Johann von Hase, Stefan Stein, Sarah Vincent, Kelly R Molloy, Gregor Kreth, et al. Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. *The Journal of cell biology*, 174(1):27–38, 2006.

[36] Leonid A Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research*, 19(1):37–51, 2011.

[37] Jesse R Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.

[38] Kashif Ahmed, Hesam Dehghani, Peter Rugg-Gunn, Eden Fussner, Janet Rossant, and David P Bazett-Jones. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS One*, 5(5):e10531, 2010.

[39] Emanuela V Volpi, Edith Chevret, Tania Jones, Radost Vatcheva, Jill Williamson, Stephan Beck, R Duncan Campbell, Michelle Goldsworthy, Stephen H Powis, Jiannis Ragoussis, et al. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *Journal of cell science*, 113(9):1565–1576, 2000.

[40] Ruth RE Williams, Simon Broad, Denise Sheer, and Jiannis Ragoussis. Subchromosomal positioning of the epidermal differentiation complex (edc) in keratinocyte and lymphoblast interphase nuclei. *Experimental cell research*, 272(2):163–175, 2002.

[41] Séverine Chambeyron and Wendy A Bickmore. Chromatin decondensation and nuclear reorganization of the hoxb locus upon induction of transcription. *Genes & development*, 18(10):1119–1130, 2004.

[42] Heiner Albiez, Marion Cremer, Cinzia Tiberi, Lorella Vecchio, Lothar Schermelleh, Sandra Dittrich, Katrin Küpper, Boris Joffe, Tobias Thormeyer, Johann von Hase, et al. Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks. *Chromosome Research*, 14 (7):707–733, 2006.

[43] Miguel R Branco and Ana Pombo. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS biology*, 4(5):e138, 2006.

[44] G Kreth, J Finsterle, J Von Hase, M Cremer, and C Cremer. Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code. *Biophysical journal*, 86(5):2803, 2004.

[45] Elphège P Nora, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L van Berkum, Johannes Meisig, John Sedat, et al. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381–385, 2012.

[46] Hiroki Yokota, Ger van den Engh, John E Hearst, Rainer K Sachs, and Barbara J Trask. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human g0/g1 interphase nucleus. *The Journal of cell biology*, 130(6):1239–1249, 1995.

[47] Manfred Bohn and Dieter W Heermann. Diffusion-driven looping provides a consistent framework for chromatin organization. *PLoS One*, 5(8):e12218, 2010.

[48] RK Sachs, G Van Den Engh, B Trask, H Yokota, and JE Hearst. A random-walk/giant-loop model for interphase chromosomes. *Proceedings of the National Academy of Sciences*, 92(7):2710–2714, 1995.

[49] Manfred Bohn, Dieter W Heermann, and Roel van Driel. Random loop model for long polymers. *Physical Review E*, 76(5):051805, 2007.

[50] Kurt Binder. Applications of monte carlo methods to statistical physics. *Reports on Progress in Physics*, 60(5):487, 1997.

[51] James D Watson et al. Molecular biology of the gene. *Molecular biology of the gene.*, (5nd edn), 2003.

[52] Jonathan R Chubb, Shelagh Boyle, Paul Perry, Wendy A Bickmore, et al. Chromatin motion is constrained by association with nuclear compartments in human cells. *Current Biology*, 12(6):439–445, 2002.

[53] LD Landau and EM Lifshitz. Statistical physics, part 1, vol. 5. *Course of theoretical physics*, page 366, 1980.

[54] Mary E Donohoe, Li-Feng Zhang, Na Xu, Yang Shi, and Jeannie T Lee. Identification of a ctcf cofactor, yy1, for the x chromosome binary switch. *Molecular cell*, 25 (1):43–56, 2007.

[55] Pablo Navarro, Andrew Oldfield, Julie Legoupi, Nicola Festuccia, Agnès Dubois, Mikael Attia, Jon Schoorlemmer, Claire Rougeulle, Ian Chambers, and Philip Avner. Molecular coupling of tsix regulation and pluripotency. *Nature*, 468(7322): 457–460, 2010.

[56] Iris Jonkers, Tahsin Stefan Barakat, Eskeatnaf Mulugeta Achame, Kim Monkhorst, Annegien Kenter, Eveline Rentmeester, Frank Grosveld, J Anton Grootegoed, and Joost Gribnau. Rnf12 is an x-encoded dose-dependent activator of x chromosome inactivation. *Cell*, 139(5):999–1011, 2009.

[57] Corinne Chureau, Sophie Chantalat, Antonio Romito, Angélique Galvani, Laurent Duret, Philip Avner, and Claire Rougeulle. Ftx is a non-coding rna which affects xist expression and chromatin structure within the x-inactivation center region. *Human molecular genetics*, 20(4):705–718, 2011.

[58] Pablo Navarro, Sylvain Pichard, Constance Ciaudo, Philip Avner, and Claire Rougeulle. Tsix transcription across the xist gene alters chromatin conformation without affecting xist transcription: implications for x-chromosome inactivation. *Genes & development*, 19(12):1474–1484, 2005.

[59] Takashi Sado, Yuko Hoki, and Hiroyuki Sasaki. ¡ i¿ tsix¡/i¿ silences¡ i¿ xist¡/i¿ through modification of chromatin structure. *Developmental cell*, 9(1):159–165, 2005.

[60] Pablo Navarro, Ian Chambers, Violetta Karwacki-Neisius, Corinne Chureau, Céline Morey, Claire Rougeulle, and Philip Avner. Molecular coupling of xist regulation and pluripotency. *Science Signaling*, 321(5896):1693, 2008.

[61] Mary E Donohoe, Susana S Silva, Stefan F Pinter, Na Xu, and Jeannie T Lee. The pluripotency factor oct4 interacts with ctcf and also controls x-chromosome pairing and counting. *Nature*, 460(7251):128–132, 2009.

[62] Erik Splinter, Elzo de Wit, Elphège P Nora, Petra Klous, Harmen JG van de Werken, Yun Zhu, Lucas JT Kaaij, Wilfred van IJcken, Joost Gribnau, Edith Heard, et al. The inactive x chromosome adopts a unique three-dimensional conformation that is dependent on xist rna. *Genes & development*, 25(13):1371–1383, 2011.

[63] Julie Chaumeil, Patricia Le Baccon, Anton Wutz, and Edith Heard. A novel role for xist rna in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes & development*, 20(16):2223–2237, 2006.

[64] Elena M Pugacheva, Vijay Kumar Tiwari, Ziedulla Abdullaev, Alexander A Vostrov, Patrick T Flanagan, Wolfgang W Quitschke, Dmitri I Loukinov, Rolf Ohlsson, and Victor 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. *Human molecular genetics*, 14(7):953–965, 2005.

[65] Jeannie T Lee. Regulation of x-chromosome counting by tsix and xite sequences. *Science*, 309(5735):768–771, 2005.

[66] S Augui, GJ Filion, S Huart, E Nora, M Guggiari, M Maresca, AF 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, 2007.

[67] Joshua Starmer and Terry Magnuson. A new model for random x chromosome inactivation. *Development*, 136(1):1–10, 2009.

[68] Masao Doi. *The theory of polymer dynamics*, volume 73. Oxford University Press on Demand, 1988.

[69] Wolfgang W Quitschke, Michael J Taheny, Laura J Fochtmann, and Alexander A Vostrov. Differential effect of zinc finger deletions on the binding of ctcf to the promoter of the amyloid precursor protein gene. *Nucleic acids research*, 28(17): 3370–3378, 2000.

[70] Ikuhiro Okamoto, Catherine Patrat, Dominique Thépot, Nathalie Peynot, Patricia Fauque, Nathalie Daniel, Patricia Diabangouaya, Jean-Philippe Wolf, Jean-Paul Renard, Véronique Duranthon, et al. Eutherian mammals use diverse strategies to initiate x-chromosome inactivation during development. *Nature*, 472(7343):370–374, 2011.

[71] IM van den Berg, RJ Galjaard, JSE Laven, and JH van Doorninck. Xci in preimplantation mouse and human embryos: first there is remodelling. *Human genetics*, 130(2):203–215, 2011.

[72] Philippe Clerc and Philip Avner. Role of the region 3 to xist exon 6 in the counting process of x-chromosome inactivation. *Nature genetics*, 19(3):249–253, 1998.

[73] Sébastien Vigneau, Sandrine Augui, Pablo Navarro, Philip Avner, and Philippe Clerc. An essential role for the dxpas34 tandem repeat and tsix transcription in

the counting process of x chromosome inactivation. *Proceedings of the National Academy of Sciences*, 103(19):7390–7395, 2006.

[74] Jeannie T Lee and Naifang Lu. Targeted mutagenesis of¡ i¿ tsix¡/i¿ leads to nonrandom x inactivation. *Cell*, 99(1):47–57, 1999.

[75] Yuya Ogawa and Jeannie T Lee. ¡ i¿ xite¡/i¿, x-inactivation intergenic transcription elements that regulate the probability of choice. *Molecular cell*, 11(3):731–743, 2003.

[76] Edith Heard, Fabien Mongelard, Danielle Arnaud, and Philip Avner. Xist yeast artificial chromosome transgenes function as x-inactivation centers only in multicopy arrays and not as single copies. *Molecular and cellular biology*, 19(4):3156–3166, 1999.

[77] Graeme D Penny, Graham F Kay, Steven A Sheardown, Sohaila Rastan, Neil Brockdorff, et al. Requirement for xist in x chromosome inactivation. *Nature*, 379(6561): 131, 1996.

[78] York Marahrens, Barbara Panning, Jessica Dausman, William Strauss, and Rudolf Jaenisch. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes & Development*, 11(2):156–166, 1997.

[79] Peter Fraser and Wendy Bickmore. Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447(7143):413–417, 2007.

[80] Alexander Gimelbrant, John N Hutchinson, Benjamin R Thompson, and Andrew Chess. Widespread monoallelic expression on human autosomes. *Science*, 318(5853): 1136–1140, 2007.

[81] Philippe Clerc and Philip Avner. Multiple elements within the¡ i¿ xic¡/i¿ regulate random x inactivation in mice. In *Seminars in cell & developmental biology*, volume 14, pages 85–92. Elsevier, 2003.

[82] Céline Morey, Pablo Navarro, Emmanuel Debrand, Philip Avner, Claire Rougeulle, and Philippe Clerc. The region 3 to xist mediates x chromosome counting and h3 lys-4 dimethylation within the xist gene. *The EMBO journal*, 23(3):594–604, 2004.

[83] Céline Morey, Danielle Arnaud, Philip Avner, and Philippe Clerc. Tsix-mediated repression of xist accumulation is not sufficient for normal random x inactivation. *Human molecular genetics*, 10(13):1403–1411, 2001.

[84] Antonio Scialdone and Mario Nicodemi. Mechanics and dynamics of x-chromosome pairing at x inactivation. *PLoS computational biology*, 4(12):e1000244, 2008.

[85] David Chandler. Introduction to modern statistical mechanics. *Introduction to Modern Statistical Mechanics, by David Chandler, pp. 288. Foreword by David Chandler. Oxford University Press, Sep 1987. ISBN-10: 0195042778. ISBN-13: 9780195042771*, 1, 1987.

[86] Heidi Sutherland and Wendy A Bickmore. Transcription factories: gene expression in unions? *Nature Reviews Genetics*, 10(7):457–466, 2009.

[87] Ben Langmead, Cole Trapnell, Mihai Pop, Steven L Salzberg, et al. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol*, 10(3):R25, 2009.