How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes

Surya K. Ghosh[1] **and Daniel Jost**[1,*]

[1]Univ Grenoble Alpes, CNRS, TIMC-IMAG, F-38000 Grenoble, France
[*]daniel.jost@univ-grenoble-alpes.fr

ABSTRACT

The 3D organization of chromosome is crucial for regulating gene expression and cell function. Many experimental and polymer modeling efforts are dedicated to deciphering the mechanistic principles behind chromosome folding. Chromosomes are long and densely packed - topologically constrained - polymers. The main challenge is therefore to develop adequate models and simulation methods to investigate properly the multi spatio-temporal scales of such macromolecules. Here, we develop a generic strategy to develop efficient coarse-grained models for self-avoiding polymers. Accounting accurately for the polymer entanglement length and the volumic density, we show that our simulation scheme not only captures the steady-state structural and dynamical properties of the system but also tracks the same dynamics at different coarse-graining. This strategy allows a strong power-law gain in numerical efficiency and offers a systematic way to define reliable coarse-grained null models for chromosomes and to go beyond the current limitations by studying long chromosomes during an extended time period with good statistics. We use our formalism to investigate in details the time evolution of the 3D organization of chromosome 3R (20 Mbp) in drosophila during one cell cycle (20 hours). We show that a combination of our coarse-graining strategy with a one-parameter block copolymer model integrating epigenomic-driven interactions quantitatively reproduce experimental data at the chromosome-scale and predict that chromatin motion is very dynamic during the cell cycle.

1 Introduction

Though all cells of a multicellular organism contain the same genetic information, they vary widely in shapes, in physiologies, and in functions. These differences mainly reflect variations in gene expression between different tissues or cell types. Recent experiments have highlighted the important role of the physical organization of chromosomes inside the cell nucleus in regulating gene expression[1–3]: gene activities being modulated, not only by the local folding of the chromatin fiber but also by its higher order organization with 3D nuclear compartments favorable to gene activation or repression. During interphase, the longest phase of the cell cycle where genes are expressed and DNA is replicated, chromosomes are found to be organized hierarchically. Confocal and electron microscopy experiments have revealed that each chromosome occupies it own territory[4]. Also, the genes sharing the same transcriptional state tend to colocalize[5–7]: inactive genomic regions (the heterochromatin) being more peripheral while active regions (the euchromatin) being more central. At the sub-chromosomal level, advanced molecular biology tools, like chromosome conformation capture (HiC) techniques, have shown that chromosomes are partitioned into consecutive 3D interaction compartments, the so-called topologically-associated domains (TADs),[8–10]. Loci inside these domains experience enriched contact probabilities with other loci of the same domain while showing partial insulation from loci of nearest neighbor domains. These domains can be easily visualized as consecutive "squares" along the diagonal of a 2D contact frequency matrix (see Fig.6 for an illustration). TADs formation has been associated with the local biochemical composition of chromatin, the so-called epigenome, which encodes for gene activity[7,11–14]: genes inside the same TAD tends to have the same epigenomic state, and long-range contacts may be observed between TADs of the same state.

However, how genome precisely organized in space is still not fully understood and addressing this question represents one of the most exciting challenges of modern biology[15]. Lots of experimental and modeling efforts are currently dedicated to understanding the mechanisms implied in chromosome folding. In particular, polymer models have been instrumental in simulating and testing different molecular and physical mechanisms and in driving new experiments[5,16–37]. An important challenge for such models is to simulate with a good precision the behavior of long polymer chains (the typical size of a chromosome ranging from about a million base-pairs in yeast to hundreds of Mbp in human) during an extended time period (of the order of hours for a typical cell cycle). Therefore the standard strategy used in these approaches is to start from a *coarse-grained "null"* model for chromatin with few basic interactions[38]. Then eventually decorate it with more physical or chemical interactions driven by biological information such as the gene activity or the local epigenomic state.

Chromosomes being very long polymers densely packed into the cell nucleus, topological constraints generated by polymer entanglement may play an important role in controlling the dynamics and fluctuations of such polymers[16,39]. However, when building their null models, very few approaches account adequately for such considerations. As neglecting topological constraints may lead to different structural and dynamical properties of the polymer, how can one interpret the outcomes correctly of such models concerning realistic mechanisms, if the considered null model is already biased?

Here, we develop a generic coarse-graining strategy for self-avoiding polymers that allows, simultaneously, to drastically reduce the computation time while maintaining the polymer in the same topological regime and thus preserving the correct structural and dynamical properties. In the first part, we explain the strategy and demonstrate its efficiency, leading to a systematic approach to developing a coarse-grained null model for chromosomes. In the second part, we apply it to investigate the role of epigenomic-driven interactions in the folding and dynamics of an entire drosophila chromosome. Finally, we discuss our results and their implications in the general context of chromosome modeling and chromatin biology.

2 Results

2.1 Chromosome and entanglement length

Chromosomes are long polymers confined inside a small volume, the cell nucleus[40]. As a result, the generic characteristics of these densely packed long polymers are very different from free isolated chains and exhibit distinct universal properties[39,41]. For a simple semi-flexible self-avoiding polymer, composed of N beads of size b, such properties are mainly determined by (i) its rigidity, characterized by its Kuhn length l_k, and (ii) by its volumic density ρ. Moreover, in a non-dilute environment, topological constraints are also expected to influence the large-scale organization and long-time dynamics. Their importance depends on the ratio between $L \equiv Nb$ and L_e, the so-called entanglement length. L_e measures the typical subchain size above which topological confinement due to excluded volume influences configurational fluctuations, and depends on l_k and ρ. It may be associated with the tube diameter in the reptation model or to the crossover time between a Rouse-like motion and a reptation-like motion[42], and may be estimated using the phenomenological relation[43]

$$L_e = l_k \left(\frac{c}{\rho_k l_k^3} \right)^2, \tag{1}$$

with $c \approx 19$ is a numerical factor and $\rho_k = \rho b / l_k$ the volumic density in Kuhn segment. In the following, we define $N_k \equiv l_k / b$ as the Kuhn segment size. $L/L_e \ll 1$ means very weak topological effects and the polymer behaves as a standard Rouse chain. If $L/L_e \gg 1$, the chain motion is restricted due to strong topological interactions and exploration of the available space is very slow. In this case, equilibration time of the chain scales as N^3[44], implying that polymer dynamics remains out-of-equilibrium and the initial topological properties (presence/absence of knots) or large-scale organization features are maintained over a long time period.

2.1.1 A reference model for chromosome

To provide the physically realistic scenario of chromosome structure and dynamics, we need to precise the values of l_k and b. We define the fine scale null model of chromosomes with Kuhn length $l_k = l_k^0$, number of beads $N = N_0$ and bead size $b = b_0$ as our reference model. Precise measurements of the Kuhn length of *in vivo* chromatin are still lacking and controversy still exists about its value, going from few nanometers[45], the so-called 10 *nm*-fiber, to hundreds of nanometers, the so-called 30 *nm*-fiber[46]. We decide to base our approach at the nucleosomal scale (1 monomer=200 bp, $b_0 \approx 10$ *nm*) with a recent estimation of $l_k^0/b_0 \approx 1$ kbp based on cyclization probabilities of chromatin[26].

In order to determine the value of the entanglement length L_e, the other important quantity to fix is the chromatin volumic density ρ. Depending on the species and on the cell type, it may strongly vary. For example, in yeast nuclei, $\rho \approx 0.005\, bp/nm^3$[47]. For drosophila, $\rho \approx 0.009\, bp/nm^3$, while being higher in typical mammalian nuclei ($\rho \approx 0.015\, bp/nm^3$). Systems with higher volumic density become more entangled and exhibit shorter entanglement length (Eq. 1). This leads to decreasing values for $N_e \equiv L_e/b_0 \approx 920$ kbp (yeast), 285 kbp (drosophila) or 102 kbp (mammals).

In yeast, $L/b_0 \approx 750$ kbp, implying that chromosomes are weakly entangled ($L/L_e \lesssim 1$). In higher eukaryotes, as drosophila or mammals, chromosomes are longer (tens or hundreds of Mbp) and in the regime of strong topological constraints ($L/L_e \gg 1$). Note that the 30 *nm* fiber model for chromatin ($l_k^0 = 300$ *nm*, $b_0 = 30$ *nm*, $l_k^0/b_0 = 30$ kbp) would lead to similar order of magnitude for L/L_e.

2.1.2 Generic behavior of chromosome

To illustrate the generic behavior of our chosen reference model in the different regimes, we perform kinetic Monte-Carlo (KMC) simulations of the chain dynamics using a lattice model (see Materials and Methods).

We study the generic behavior of the reference model in the "yeast" ($N_0 = 750$ kbp $= 3750$ monomers, $\Phi_0 = 0.023$) and in the "drosophila" ($N_0 = 20$ Mbp $= 10^5$ monomers, $\Phi_0 = 0.043$) cases. The bending rigidity associated to the reference Kuhn length for both cases ($l_k^0/b_0 = 1$ kbp $= 5$ monomers) is $\kappa_0 = 3.12$ (in k_BT-unit).

During the equilibration of simulation chains, we measured all the standard physical quantities: $g_1(t)$, $g_3(t)$, $\langle R^2(s)\rangle$ and $P_c(s)$ (Fig.1). Comparison of g_1 with 0.01 $t^{0.5}$, the experimentally measured typical value of g_1 in μm^2 for yeast and drosophila[45,48], gave the equivalency of each *MCS* with real time in *sec*. From the time mapping we were able to represent our results in real physical unit: time in *sec*, distance in μm; the 10^7-MCS long trajectories corresponding to about 30 *min* of real time.

The yeast chromosome behaves dynamically as a standard Rouse chain[45]. At intermediate times, $g_1 \sim t^{0.5}$ (Fig.1a), the typical scaling law in the Rouse diffusion regime[49]. At later time, when t is greater than the Rouse time, the typical time by which the polymer has already traveled a distance equivalent to its size, g_1 coincides with the center of mass MSD ($g_3 \sim t$), characteristic of a simple diffusion process[49]. In the drosophila case, topological constraints are strong and the anomalous diffusion exponents of g_1 at intermediate time scale behaves as $t^{0.4}$ (Fig.1a,d). We demonstrated how the scaling behavior of MSD changes as we vary the measurement time window and reached the steady state in Fig.S2. Note that, we do not observe the scaling exponent, at least in the scanned time scales, expected from reptation dynamics ($t^{0.25}$) of entangled polymers[44]. This is a characteristic of knot-free polymers, like crumpled or ring polymers[50] and is reminiscent of the initial unknotted configurations. We verified our simulation scheme for different initial configurations: hedgehog, rosette, cylindrical, knotted, and random configurations (Fig.S3), we concluded that our simulation strategy does not depend on the type of initial conditions. Starting from random configurations that contain knots, we recover the standard reptation regime (Fig.S4a,d). We have also constructed initial configuration with a fixed number of knots. Presence of few number of knots has very little effect on the structural and dynamical properties (Fig.S5). In the yeast and drosophila cases, at small time scales ($t <$ms), g_1 scales as $t^{0.75}$ which corresponds to the typical diffusion regime of a semi-flexible chain up to the Kuhn length scale[51].

Regarding the structural properties $\langle R^2(s)\rangle$ (Fig.1b) and $P_c(s)$ (Fig.1c), we recover the main scaling laws observed experimentally for chromosomes of yeast, fly and other eukaryotes[5,6,41,52–54]. The yeast case is fully consistent with a worm-like-chain at equilibrium with $\langle R^2(s)\rangle \sim s^1$ and $P_c \sim s^{-1.5}$[49]. On the other hand, for the fly chromosome, the scaling laws ($\langle R^2(s)\rangle \sim s^{2/3}$ and $P_c \sim s^{-1.1}$) are consistent with crumpled polymer physics[16,41,55–57]. The large scale behavior ($s > 1$ Mbp) is a remaining signature of the initial scaling laws (Fig.S6b, S7b): the system has yet to reach steady-state and is still strongly out-of-equilibrium.

2.2 Coarse-graining long polymers at fixed entanglement length

Using the fine scale reference model, we recover the expected structural and dynamical behaviors in the different entanglement regimes, fully consistent with previous theoretical works on knot-free and crumpled polymers[16,41,50,55,56,58] and with experiments[5,6,45,48,52–54,59]. At this nucleosomal scale, the model has very good spatial (10 *nm*) and temporal (15 μsec, 30 *min* $\approx 10^8$ MCS) resolutions. However, the underlying cost of this is a prohibitive computational time. For example, for long chromosomes such as in the drosophila case, simulating one 30 *min* long trajectory requires 83 hrs CPU time on a 3.20 GHz CPU. To access more biologically-relevant time-scales (dozens of hours) with good statistics, we aim to develop a coarse-graining strategy of the reference model that allows to speed up the simulation of long trajectories while preserving the main physical characteristics of the original -fine scale - model.

2.2.1 Coarse-graining strategy

We consider an arbitrary fine-scale model (FSM) of a semi-flexible self-avoiding walk defined by N_0, l_k^0, b_0. We note N, l_k and b, the corresponding values of a coarse-grained model (CGM) of the FSM. Each CGM monomer encompasses $n = N_0/N > 1$ FSM monomers. A possible CG strategy consists in neglecting the bending rigidity ($l_k = b$) in the CGM if n is greater than the Kuhn size $N_k^0 \equiv l_k^0/b_0$, in the FSM, and in imposing the size of CGM monomer to equal the mean end-to-end distance of the corresponding number of FSM monomers, ie $b = \sqrt{nb_0 l_k^0}$. Using Eq.1 and conservation of total volume, it is easy to show that the ratio (L/L_e) is also conserved, and therefore the effect of topological constraints. However, such approach is limited by the volume fraction Φ occupied by the CG chain. Indeed, for lattice or off-lattice (assuming spherical shape for monomer in the FSM and CGM) models

$$\frac{\Phi}{\Phi_0} = n^{1/2}\left(\frac{l_k^0}{b_0}\right)^{3/2} > 1 \tag{2}$$

Hence, if Φ_0 is already high in the FSM and/or the coarse-graining is strong ($n \gg 1$), Φ might become close or higher to 1 and therefore very hard to simulate. For example in the case of drosophila chromosomes ($\Phi_0 \sim 4.3\%$), if $n = 5$ (corresponding to 1 kbp resolution, the Kuhn segment size of the FSM), $\Phi = 25\Phi_0 > 1$. Therefore, already at the Kuhn size scale, such CG strategy may fail to generate simulable models. Forcing the CG to high n anyway would imply to consider smaller b, risking

not to well account for distances, and also to decrease the ratio L/L_e. This may affect the dynamical regime of the chain, and hence its physical properties (see Section 2.3). To avoid this, we develop a novel coarse-graining strategy that allows to go for high coarse-graining while keeping the volumic fraction in a simulable range and the ratio (L/L_e) fixed.

We authorized the CGM, even if $n > (l_k^0/b^0)$, to have a bending rigidity characterized by l_k. The conditions we imposed here is that the ratio L/L_e and the volume of the simulation box are conserved. In the lattice framework, using Eq. 1, these constraints can be reformulated as (see Materials and Methods)

$$\frac{l_k}{b} = \left[\left(\frac{b_0}{L_e^0}\right)\left(\frac{c^2}{2n\Phi^2}\right)\right]^{1/3} \quad \text{and} \quad b = \left(\frac{\sqrt{2}n\Phi}{\rho}\right)^{1/3} \tag{3}$$

with $\rho \equiv N_0/V$ the volumic density in FS monomers (with V the volume of the box). Φ now plays the role of a control parameter and can be arbitrarily chosen: the characteristics of the CGM depends not only on the FS properties but also on Φ (see Table 1 for examples): a given Φ determines b and l_k, and the corresponding value for the lattice bending energy κ is inferred from l_k/b (see Materials and Methods). It has to be noted that the corresponding CG bending rigidity is not a "true" rigidity that reflects the rigidity of the FSM. It is an artificial rigidity, allowing to control Φ. Therefore, the CGM cannot pretend to quantitatively describe the FSM properties at scales smaller than few l_k.

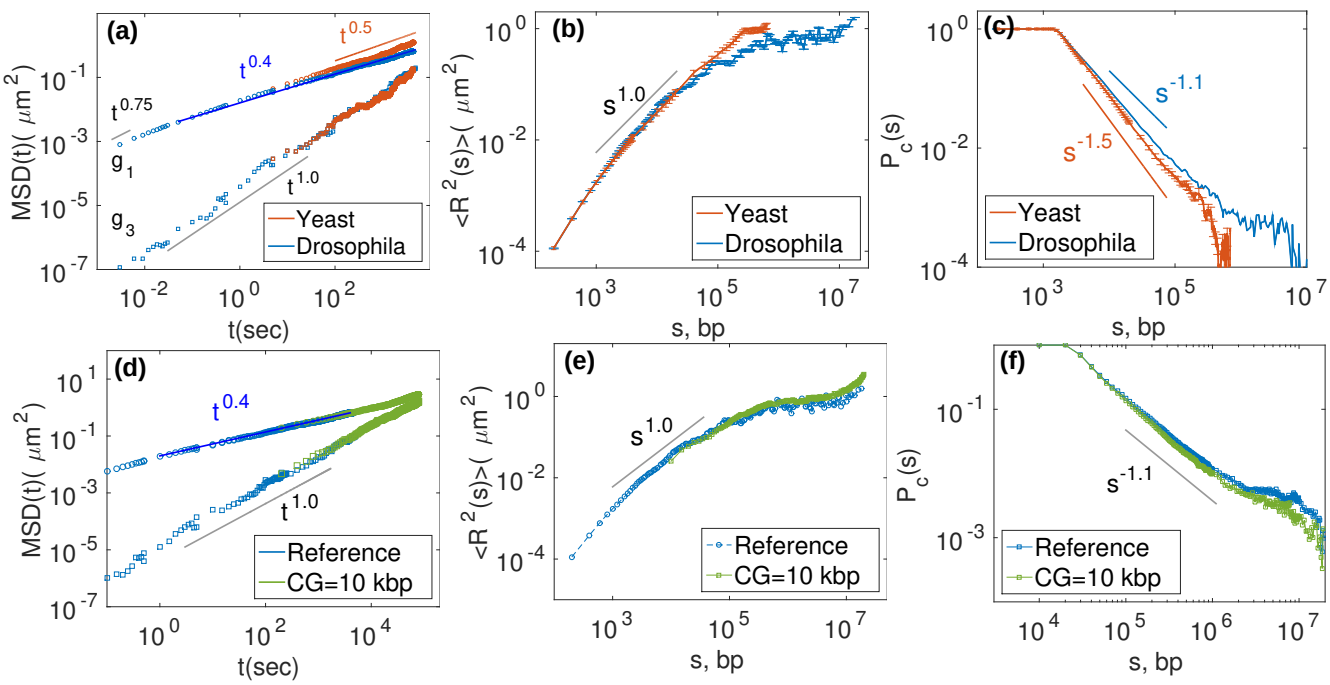

Figure 1. In the top panels, we compare different physical properties for yeast (red) and drosophila (blue) chromosomes at a nucleosomal resolution of 200 bp (reference model). In the bottom panel, we compare the reference model ($\Phi_0 = 0.043$) with one possible coarse-grained model ($CG = 10$ kbp, $\Phi = 0.97$) for drosophila chromosome. (a,d) Individual MSD $g_1(t)$ (top curves), and center of mass MSD $g_3(t)$ (bottom curves) as a function of time. (b,e) The average physical squared distance $\langle R^2(s)\rangle$ between any two monomers as a function of their linear distance s along the chain (given in bp). (c,f) Average contact probability $P_c(s)$ between any two monomers as a function of their linear distance s along the chain (given in bp) (with $d_c = 55$ nm in (c) and $d_c = 163$ nm in (f)). In (e,f), averages were computed over the same real time window (100 $sec - 30$ min). The error bars in (a, b, c) were computed as the standard deviation of the mean. Error bars in (a) are of the similar size of the symbols.

2.2.2 Conservation of generic properties and time mapping

In this part, we test if the above coarse-graining strategy conserves the structural and dynamical properties of the reference fine-scale model. In this article, we primarily focus on the drosophila case. However, in the Supplementary Information, we show that the method also performs very well for the yeast and mammalian cases (Fig.S8, S9) and that the success of the strategy does not depend on the type of used initial conditions (Fig.S3).

Table 1. Simulation parameters for the drosophila 3R chromosome.

200 bp					2 kbp					5 kbp					10 kbp				
Φ_0	N_k^0	b_0	l_k^0	*MCS*	Φ	N_k	b	l_k	*MCS*	Φ	N_k	b	l_k	*MCS*	Φ	N_k	b	l_k	*MCS*
0.043	1.0	10.6	55.4	0.015	0.027	29	20.5	305	0.002	0.09	44	41.9	374	0.4	0.37	44	83.8	372	8.71
					0.039	23	23.1	273	0.010	0.17	29	51.3	304	1.2	0.68	29	102.5	304	17.3
					0.049	20	24.9	252	0.040	0.24	23	57.6	271	1.7	0.97	23	115.3	274	22.1
					0.076	15	28.8	218	0.083	0.29	20	61.5	252	2.5					
					0.113	11	32.9	190	0.181	0.45	15	70.9	220	5.4					
					0.113	11	32.9	190	0.181	0.45	15	70.9	220	5.4					

Table for simulation parameters and time mappings used for the drosophila 3R chromosome, at different coarse-graining: lattice volumic fraction Φ, Kuhn size $N_k \equiv l_k/b$ (in kbp-unit), bond length of the polymer b (in *nm*) and Kuhn length l_k (in *nm*). The corresponding simulation parameters associated to the fine-scale reference model are Φ_0, N_k^0, b_0 and l_k^0. Each fifth subcolumn represents the time in *msec* equivalent to one Monte-Carlo steps (MCS). Similar tables for yeast and mammal cases are given in the supplementary text.

In Fig.1 bottom panel, we compared results between the FSM and a CGM at 10 kbp resolution for $\Phi = 0.97$. Like for the reference model (see Sec.2.1.2), we time-mapped the simulation time for the CGM using $g_1(t)$ in order to have a direct time correspondence between the FSM and CGM (see Table 1). For this CG, 1 MCS represents a time step 10^4 fold larger than the FSM, meaning that the 10^8 MCS-long trajectories can span more than 100 hours (instead of 30 *min* for the FSM). We remark that MSD curves overlap nicely and that the scaling laws are conserved (Fig.1d). From the configurations collected during the "real" time window ($0-30$ *min*) common to both models, we calculated the averaged properties of $\langle R^2 \rangle(s)$ and $P_c(s)$ (Fig.1e,f). For $s < 1$ Mbp, we observe a perfect match between CG predictions and the FSM behavior. For $s > 1Mb$, the system does not reach steady-state and keeps a partial memory of the initial scaling laws that are different for both systems (Fig.S6, S7), leading to small deviations between the predictions, especially for P_c which is more sensitive to local structures.

In Fig.2, we performed similar comparisons between two different CGM at a fixed Kuhn size value $N_K \equiv l_k/b$ ($N_K = 23$ kbp): ($CG = 5$ kbp, $\Phi = 0.24$) and ($CG = 10$ kbp, $\Phi = 0.97$). As before, we recovered identical scaling laws for $g_1(t)$, the 10 kbp-resolution model allowing to scan longer times for the same number of MCS steps (Fig.2a). We computed $\langle R^2 \rangle$ and P_c for a series of snapshots taken at several real time points, at 1 *min*, 30 *min* and 10 *hrs* (Fig.2b,c). Remarkably, starting from similar behaviors for $\langle R^2(s) \rangle$ for the two CG (compare the 1 *min*-curves), the predicted time-evolution of the two curves remains identical, even after simulating more than 10h of real time. A similar comparison is also observed for $P_c(s)$.

To test that controlling the volumic fraction Φ, or equivalently the Kuhn fragment size N_K, in our strategy does not impact the coarse-graining, we perform simulations at the same CG ($= 10$ kbp), but for different values of N_K (Fig.2). Identically, we observe almost perfect matching for the time-evolution of $\langle R^2 \rangle$ and P_c for all ranges of genomic distance. To push our strategy to the limit, we also considered very high values for Φ (Fig.S10a,b,c). In our lattice polymer model, each point can be occupied by two consecutive monomers (see Sec.4.1), so in principle, a maximum volumic fraction $\Phi = 2$ is achievable. For $\Phi \lesssim 1$, all the simulations show exactly same results as the reference model and follow the same curve. For $\Phi \gtrsim 1$ the results deviate from the reference model strongly, the dynamics are dramatically slowed down due to the incapacity of the algorithm to move the monomers efficiently.

All this demonstrates that our coarse-graining strategy allows to describe the correct structural and dynamical properties of the underlying model at all scales at steady-state but more importantly also out-of-equilibrium as long as the initial configurations share the same statistical behaviors and the chosen volumic fraction is not too high. What is the gain in term of numerical efficiency? Decreasing the number of beads by augmenting the CG would automatically linearly reduce the time to compute 1 Monte-Carlo time-step in our simulations. In addition, as the CG and the controlled Φ (or equivalently N_K) are increasing, the mesh size of the lattice model (or the size of a monomer) augments, and thus 1 MCS will correspond to a larger real time step (see Tab.1). Therefore, the simulation of a fixed time period will be consequently decreased. All in all, we observed a fast polynomial decay of the numerical effort as a function of the CG scale (CPU time $\sim$ CG^{-5}, Fig.3a), with, for example, a gain of almost four orders of magnitude between the reference fine scale model and the CGM at 10 kbp resolution with $N_k = 23$ kbp (or $\Phi = 0.97$). Similarly we gain polynomially (CPU time $\sim N_k^3$) in computational time for smaller Kuhn size N_K (see Fig.3b). However smaller N_k corresponds to higher Φ which may impose restrictions on the dynamics if $\Phi > 1$.

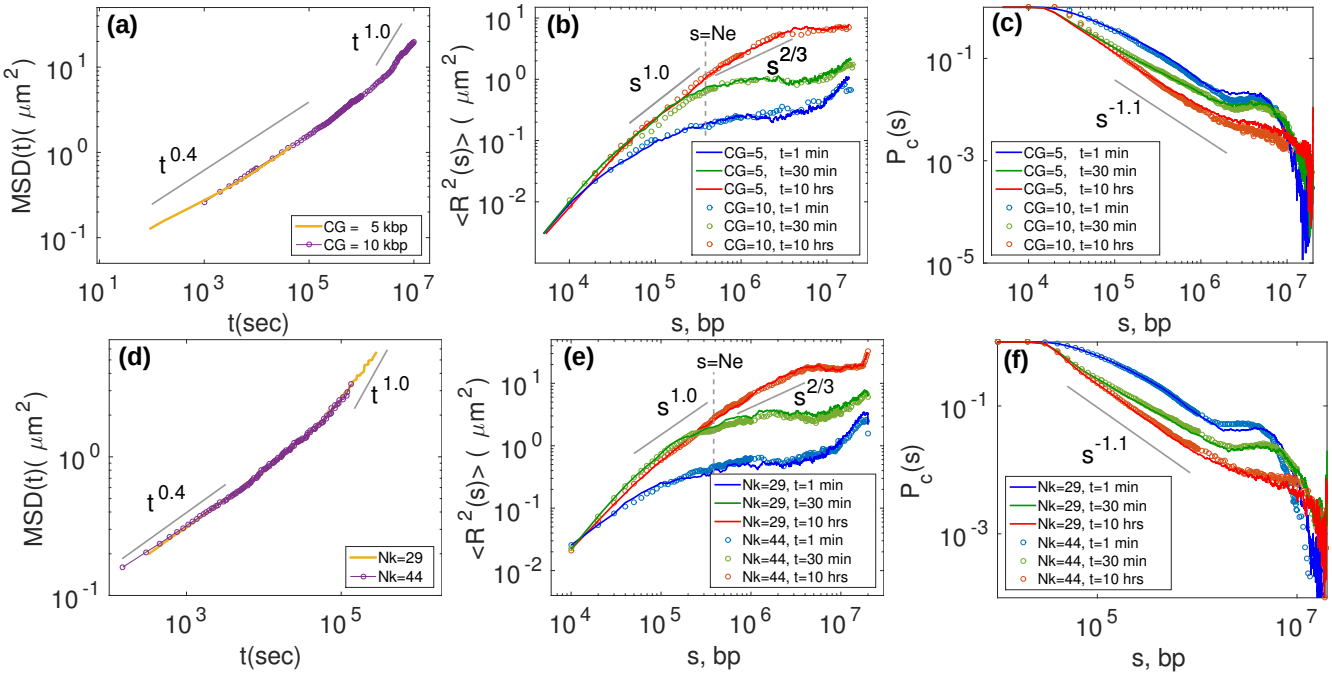

Figure 2. Comparison of physical properties and their time evolution for two different coarse-grainings ($CG = 5,\ 10$ kbp) at a fixed Kuhn size of $N_k = 23$ kbp (top panels) or for two different Kuhn sizes ($N_k = 29,\ 44$ kbp) at a fixed coarse-graining of CG=10 kbp (bottom panels). (a,d) Average MSDs as a function of time in *sec*, calculated from the trajectory of 10^7 Monte Carlo steps. Time evolution of $\langle R^2(s)\rangle$ (b,e) and $P_c(s)$ (c,f) for different coarse-grainings (b,c) and Kuhn sizes (e,f).

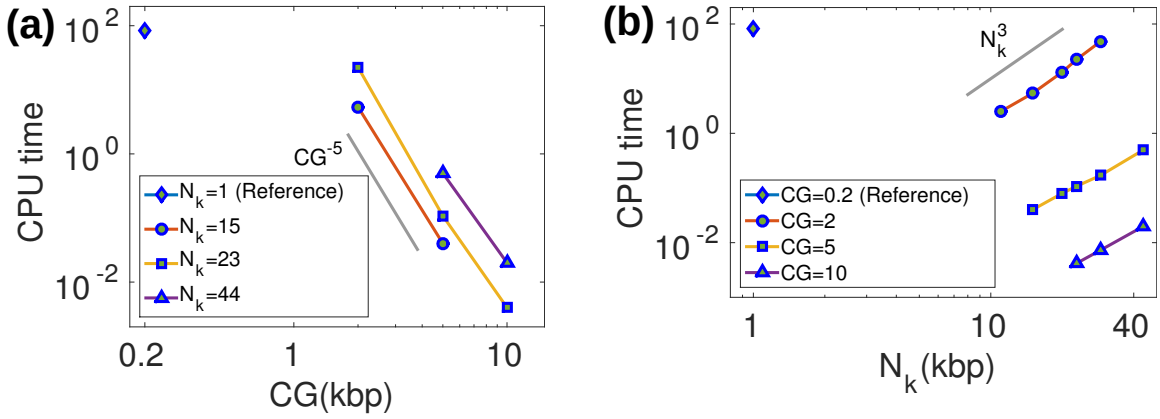

Figure 3. CPU time (given in hours) required to simulate 30 *min* of real dynamics on a 3.20 GHz processor, (a) as a function of coarse graining where Kuhn sizes are fixed at three different values and, (b) as a function of Kuhn size at three different coarse grainings. The reference model is represented by $N_k = 1$ kbp and $CG = 0.2$ kbp.

2.3 How to build a good coarse-grained null model of chromosome?

Our coarse-graining strategy is generic and has no direct connection to any particular polymer. When applied to a specific system, the question would be how to choose the optimal coarse-graining? As we observed above, regarding numerical efficiency, one wants to go for the higher CG and higher Φ ($\lesssim 1$). These two values will determine the spatial and temporal resolutions of the model. Therefore, a natural choice would be to maximize CG and Φ under the constraints of a minimal desired resolution (e.g., determined by the experimental precision). For example, for chromosomes, we aim to be quantitative typically at a 20 kbp scale ($N_K \sim 20$ kbp) with a spatial precision of about 100 *nm*. Under this loose constraints, $CG = 10$ kbp and $\Phi = 0.97$ is a very appropriate choice (see Tab.1).

Till now, throughout our study, we strategically choose the bending stiffness l_k or the lattice volumic fraction Φ so that we preserve the physical properties of the system by conserving the right entanglement length. Now the question is what happens

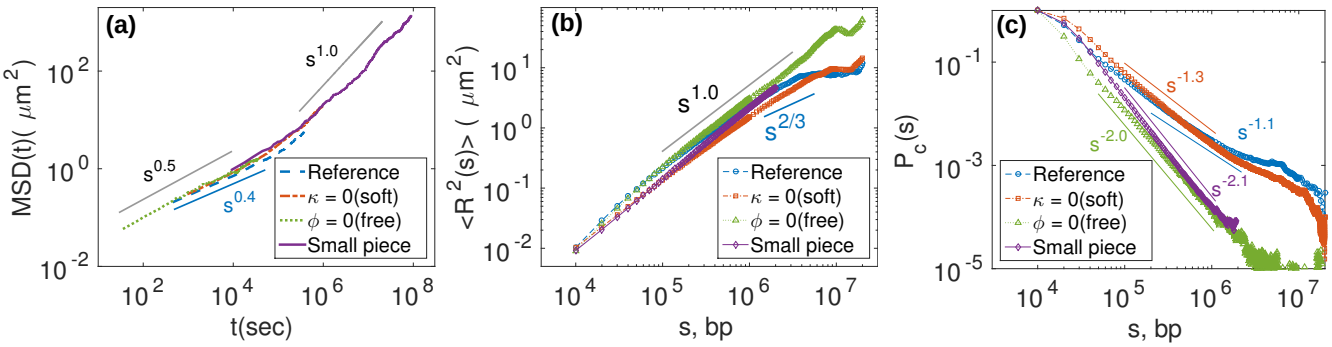

Figure 4. Comparison of reference model with different chains constructed in various extreme simulation conditions: soft $\kappa = 0$, isolated $\Phi << 1$ and small pieces with chain length less than N. (a) g_1, (b) $\langle R^2(s)\rangle$ and (c) $P_c(s)$.

if one uses more naive coarse-graining strategies that do not necessarily preserve the topological regime. As explained at the beginning of Sec.4.2, a typical strategy is to "neglect" the artificial bending rigidity if the CG is higher than the Kuhn length of the reference model. Another is to consider an isolated polymer and to neglect the "confinement" of the chain. These two kinds of models modify the L/L_e ratio and therefore might change the physical properties of the system: chain motion may be slightly accelerated ($g_1(t) \sim t^{0.5}$, Fig.4a) and structural properties may be strongly perturbed (Fig.4b,c). In particular, considering an isolated chain ($\Phi \ll 1$) dramatically modifies the behavior of P_c that scales within this approximation as $\sim s^{-2}$, characteristics of isolated self-avoiding walks[44]. This leads to an underestimation of the contact probability by orders of magnitude compared to the reference model. In the supplementary article, we did two independent studies to understand the effect of bending stiffness and lattice volume fraction where we do not restrict them to preserve the entanglement length. In the bottom panel of Fig.S10, we represented different physical properties as we vary Φ independently. We also showed the effect of bending rigidity on physical properties while keeping other simulation parameters constant in Fig.S11. Both these studies confirm that an arbitrary choice of simulation parameters could destroy the true physical properties of the system.

In complement to coarse-graining strategies, still in the purpose of reducing the computation time, a standard approach is to simulate only small pieces of the chromosomes instead of the full length. Since the dynamical regime of the chain depends on the ratio L/L_e, reducing the value of L may modify the dynamics of the chain and therefore may lead to wrong predictions. For example, instead of 20 Mbp-long polymer, if we simulate small fragments of 2 Mbp, we found strong discrepancies. At small length scales, $s < 100$ kbp, it follows the reference model, but at larger length scale 100 kbp $< s <$ 2 Mbp it deviates from reference model and behaves like an isolated Rouse polymer (see Fig.4). A systematic study of how the physical properties change as we simulate chain fragments of varying lengths is presented in Fig.S12.

All this emphasizes the need to conserve the ratio properly L/L_e of the reference model if one aims to simulate the right polymeric behavior accurately. Modifying this ratio by decreasing L or by making approximations that affect L_e would possibly lead to simulate a system with different physical properties than the actual system of interest.

2.4 Application to chromatin folding in drosophila

Having in hands a strategy to build an efficient coarse-grained "null" model for chromosome, we use it to study the folding of fly chromosome 3R. In drosophila, the 3D structural units, the so-called TADs, are strongly associated with the 1D epigenomic domains[11,60,61]: a locus of a given epigenomic state is likely to share its local 3D compartment with loci of the same epigenomic state. This observed correlation had recently motivated us to build a heteropolymer model accounting for the epigenome folding into interacting TADs[20,32]. Based on biochemical evidence that proteins associated with some epigenomic states have the capacity to oligomerize[62–64], hence possibly generating effective specific interactions between loci of the same state, we developed a block copolymer model of fly chromatin where each block represents an 1D epigenomic domain. By varying the strength of these specific interactions, we showed that such model well accounts for the TAD formation and for inter-TAD long-range contacts. Previously, we limited our analysis to short pieces of chromatin ($\sim$ 1 Mbp-long fragment) at equilibrium. In section 2.3 of the present paper, we observed that simulating only small regions instead of the full system might lead to strong approximations. Here, we wonder if our conclusions on the 3D chromosome folding in drosophila remain valid and can be generalized when considering a larger genomic region and using a more realistic "null" model for chromatin.

We consider the 20-Mbp region of chromosome 3R localized between 7 and 27 Mbp, that we model using an efficient coarse-graining (10 kbp and $\Phi = 0.97$, see Table 1). For this region, we collect the epigenomic domains obtained by Filion et al[65] for the embryonic cell line Kc167. In this dataset, five types of epigenomic states exist: 2 euchromatic states associated with active genes that, for simplicity, we decided to merge into one "active" state; and 3 heterochromatin states: constitu-

tive heterochromatin associated with HP1-protein and H3K9me3 histone marks, facultative heterochromatin associated with Polycomb (PcG) proteins and H3K27me3 histone marks, and the so-called "black" chromatin, the prevalent form of heterochromatin. To each 10-kbp monomer of the model, we associate the corresponding epigenomic state, and we assume that monomers of the same state may specifically interact with an energy E_i if they are spatially in contact (ie nearest-neighbor on the lattice). For simplicity, we assume that the strength of interaction is similar for every epigenomic state. For various values of E_i, we simulate 400 trajectories during 10^7 MCS starting from a random unknotted "hedgehog" configuration (as in Fig.S1a,c) at a standard *in vivo* bp-density ($\rho = 0.009 bp/nm^3$) (see the Movie for example).

2.4.1 Effect of varying the strength of specific interactions

We first concentrate on the polymer dynamics by studying the average MSD along the simulations (Fig.5a). For all investigated interaction energies, the scaling properties of g_1 are compatible with diffusion of a crumpled polymer as seen in Sec.2.1.2 with $g_1 \sim t^{0.35-0.4}$. As we negatively increase the interaction strength, there is a dramatic slowing-down of the polymer dynamics. Interestingly, by plotting the mean MSD at 10^6 MCS as a function of E_i in k_BT unit (Fig.5b), we observe a transition between a "fast" ($E_i > -0.25$) and a "slow" ($E_i < -0.25$) regime. This suggests a glass-like[66,67] dynamic transition that occurs for strong specific interactions, reducing significantly the monomer mobility in the simulations.

For each E_i, we performed the time mapping strategy (see above) to adjust the simulation time (MCS) to the real time. Then, we computed the average contact maps between 0 and 20 hrs ($d_c = 163\ nm$), representing the average inside a population of unsynchronized cells with a typical cell cycle of 20 hrs[47]. From our 10^7 MCS trajectories, this was possible only for E_i-values in the fast regime ($E_i \geq -0.25$). For example, for $E_i = -0.3$, the average map is only between 0 and 4h, and for $E_i = -0.4$, between 0 and 300 *sec*. At very weak interaction strengths, the polymer is crumpled as described in Sec.2.1.2. As E_i is negatively increased, blocks i.e. epigenomic domains, start to collapse forming TADs, long-range interactions between TAD of same type appear (Fig.5c) and TADs of different types segregate, which is characteristics of microphase separation in block copolymer[68]. From the contact maps, we estimate the sequence-average contact probability $P_c(s)$ as a function of the genomic distance s, as well as the average contact probability $P_{intra}(s)$ (resp. $P_{inter}(s)$) between loci of the same (resp. different) epigenomic state. As expected, stronger interactions favor (resp. unfavor) contacts at all scales between monomers of the same (resp. different) type (Fig.5d, center and right). Interestingly, while we observe opposite behaviors for P_{intra} and P_{inter}, for $0 \geq E_i \geq -0.2$, the global sequence-average probability P_c remains identical to the "null" model without interaction (Fig.5d, left), the increase in intra-state contacts being compensated by the decrease in inter-state contacts. At some point ($E_i \leq -0.2$), insulation becomes maximal and only intra-state contacts augments, leading to also an augmentation in P_c.

2.4.2 Comparison with experimental data

We next compare our results to Hi-C data obtained by Sexton et al for late drosophila embryos[60]. Experimental data exhibit the characteristic presence of TADs along the diagonal of the HiC map and of preferential long-range contacts between some TADs (Fig.6a and Fig.S13). The sequence-average probability $P_c(s)$ shows different regimes (Fig.6b): for $s < 100$ kbp, $P_c(s) \sim s^{-0.5}$, for 100 kbp $< s <$ 1 Mbp, $P_c(s) \sim s^{-1}$, for 1 Mbp $< s <$ 10 Mbp, $P_c(s) \sim s^{-0.5}$. Contacts between loci of the same epigenomic state are about 1.5-fold more pronounced at almost all scales than between loci of different states (Fig.6b).

While we do not expect our model to be quantitative at small genomic scales ($s < 100$ kbp) due to the coarse-graining we used in our simulations, the predicted shape of $P_c(s)$ is very similar to the experimental one with $P_c(s) \sim s^{-1.1}$ for 100 kbp $< s <$ 1 Mbp and $P_c(s) \sim s^{-0.5}$, for 1 Mbp $< s <$ 10 Mbp. Among the different strength of specific interactions that we investigated, $E_i = -0.1$ offers the best match with experimental data with also an enrichment of 1.5 fold of intra-state vs inter-state contact probabilities. Comparison between the predicted and the experimental contact map shows very good correlations at the local - TAD - level but also at higher scales (Fig.6c,d and Fig.S13), in terms of patterning but also in terms of relative contact frequencies. Given the simplicity of the model, it is remarkable that such model is in quantitative agreement with experimental data from small to large scales, suggesting that epigenomic-driven forces are main players of the chromosome folding in drosophila, generalizing our previous findings made on Mbp-genomic regions[20,32].

2.4.3 Dynamics of TAD formation and inter-TAD interactions

One strong conjecture of our previous study was that long-range TAD interactions are dynamical and that TADs may form very rapidly just after the mitotic exit[20]. Now that we have a more complete and largest-scale model with a proper time mapping, we aim to verify and to characterize these hypothesis. For $E_i = -0.1$, we compute the population-average contact map of synchronized cells at different times along the simulations.

Time-evolution of the predicted HiC-maps shows that TADs form very quickly in about half a minute (Fig.7a). Specific long-range contacts between monomers of the same epigenomic state are more slowly formed, ranging from minutes for sub-Mbp-scale contacts to hours at 10 Mbp-scale (Fig.7a). This is confirmed by analyzing the time-evolution of the average ratio between P_{intra} and P_c for different scales (Fig.7b). For $10-100$ kbp range, $\langle P_{intra}/P_c \rangle$ reaches a plateau after about 5 *min*, suggesting that local interactions reaches steady-state very early in the cell cycle. For the 100 kbp - 1 Mbp, convergence to

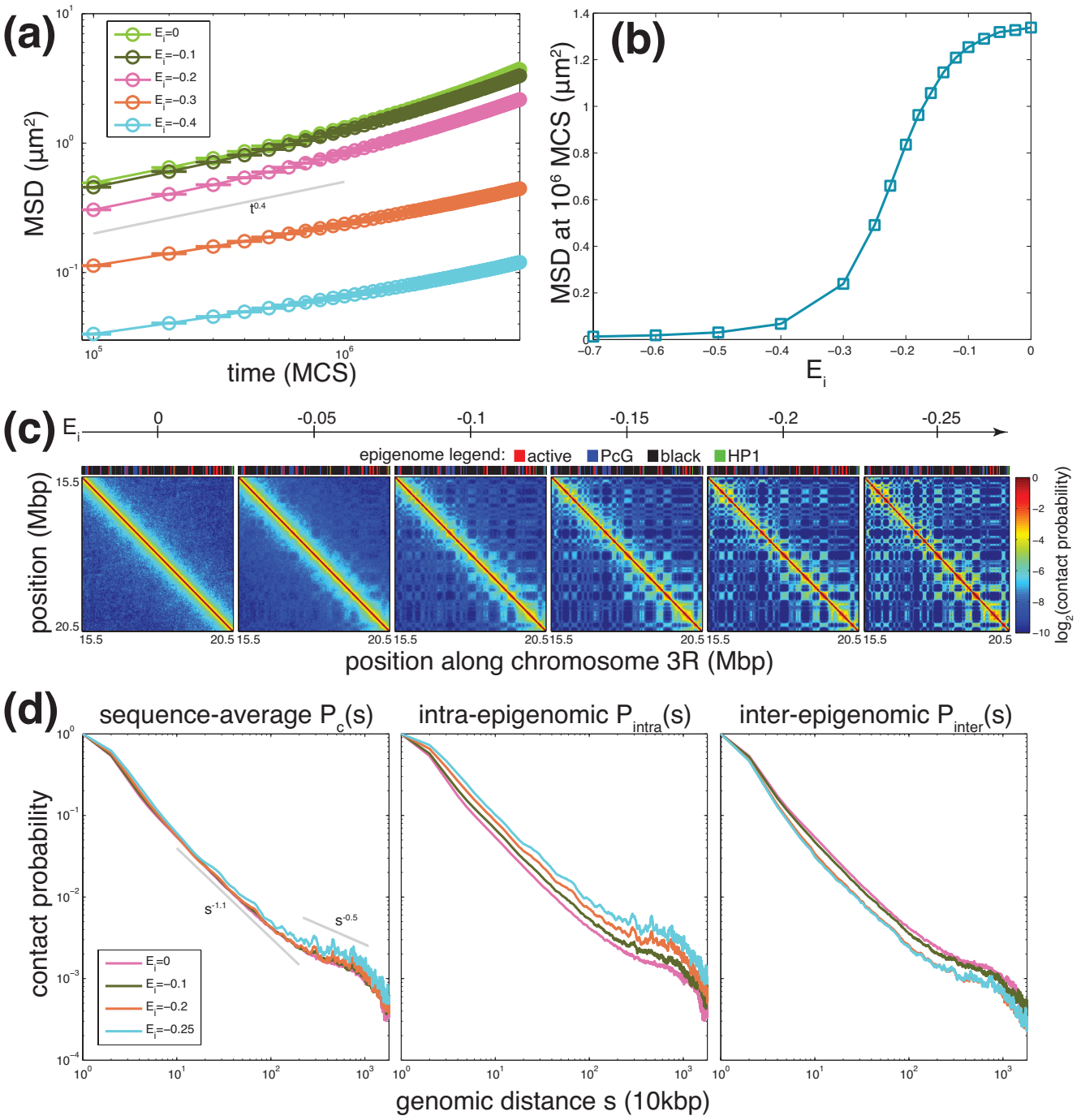

Figure 5. Dynamical and structural properties predicted by the model. (a) Average mean-squared displacement (MSD) along simulated trajectories as a function of simulation time in Monte Carlo step (MCS)-unit, for different values of the specific epigenomic-driven contact energy E_i (in k_BT-unit). (b) Average MSD after 10^6 MCS as a function of E_i. (c) Evolution of the average (between 0 and 20h) contact maps for the region located between 15.5 and 20.5 Mbp of chromosome 3R as a function of E_i. We also plot on top the underlying epigenomic landscape. (d) Average contact probability as a function of the genomic distance s between any pairs of monomers (left), between pairs of monomers having the same (center) or different (right) epigenomic state. Grey lines represents scaling laws.

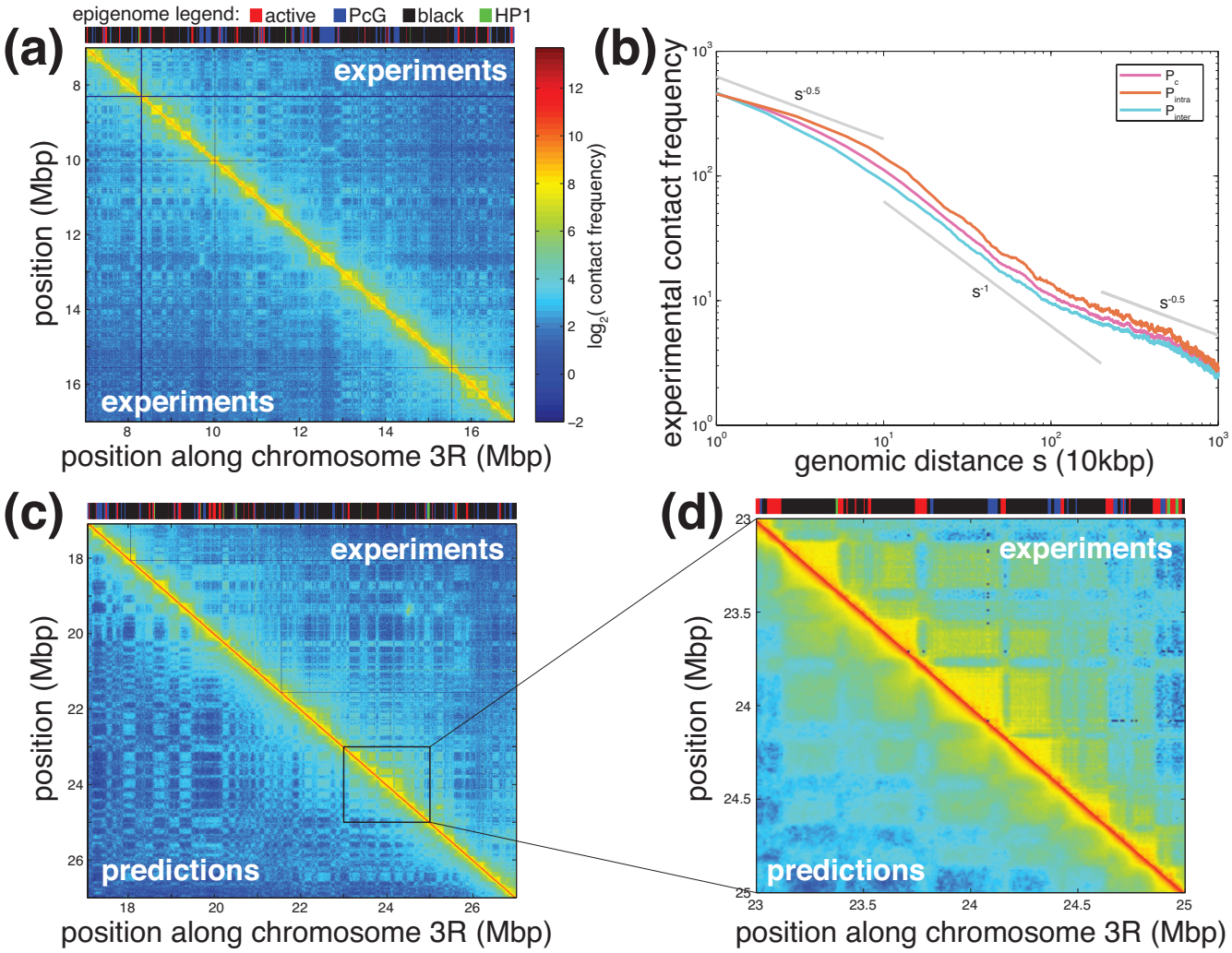

Figure 6. Comparison between experimental and predicted data. (a) Experimental Hi-C map for a 10 Mbp region of chromosome 3R. Corresponding epigenome is shown on top. (b) Average experimental contact frequency as a function of the genomic distance s between any pair of monomers (pink), between pairs of monomers having the same (orange) or different (cyan) epigenomic state. Grey lines represents scaling laws. (c, d) Predicted ($E_i = -0.1\ kT$) vs experimental contact maps for a 10 Mbp and a 2 Mbp region. Predicted data were multiplied by a factor 2500 to adjust scale with experiments.

steady-state is slower (less than 1 hour), while for longer-range interactions it takes more time (about 5h). Insulation between loci of different epigenomic states evolves at the same time-scales (Fig.7b).

To quantify the dynamics of long-range contacts between TADs, we tracked during one cell cycle (20h) with great precision (snapshots every 100 MCS ≈ 3.5 *sec*). Six pairs of loci having identical or different epigenomic states (Table S3) and separated by different genomic lengths s ($s \sim 400$ kbp, $s \sim 3$ Mbp and $s \sim 12$ Mbp). Fig.7c shows a typical time-evolution of the distance between one pair of loci in one simulated trajectory. From the trajectories, we extract three quantities: the time of first encounter τ_{first} defined as the first time after the mitotic exit when the pair becomes closer than $d = 325$ *nm*; the contact time τ_c defined as the time the pair stays in "contact" (ie closer than d); and the search time τ_s defined as the time interval between two "contacts". For each pair, the probability distribution function of τ_{first} is polynomial with two regimes with a slower decay for $\tau_{first} < 0.2$h (Fig.7d left). The scaling laws depend only on the genomic distance s between the loci, distant pairs needing more time to first contact. The polynomial dependence implies that very long τ_{first} are significantly observed. Interestingly, for $s > 1$ Mbp, it exists a significant proportion of cells (5% for $s \sim 3$ Mbp and 14% for $s \sim 12$ Mbp) where the distance between the two regions never goes below d. The distributions of τ_c are also polynomial for long times (Fig.7d center), the behavior depending only if pair members have the same or different epigenomic states. While, all the scaling laws are very similar, long contacts for pairs of loci with the same state are more frequent. In average, a contact between same-state loci lasts 12 seconds while the contact duration is divided by two for regions with different states. The distributions for τ_s are polynomial with two regimes (Fig.7d right). While the "small" time regime depends if the epigenomic states are identical or not, different-state loci being more likely to wait more between two contacts, the "long" time regime depends mainly on the genomic scale, distant loci needing more times to contact. Indeed, for short τ_s, there is still a memory of the relative positions of the two loci and pairs of the same state would be more likely to contact again rapidly, for long τ_s, memory is lost and the time interval between two contacts relies only on the genomic distance as for τ_{first}.

3 Discussion and Conclusion

In this article, in the first part, we introduced a new coarse-graining strategy for long and dense self-avoiding walks that conserves the entanglement length and the volumic density. Using kinetic Monte Carlo simulations on a lattice, we demonstrated that such strategy leads to the accurate description of, not only the steady-state but also the time-evolution of the expected structural and dynamical properties of the underlying fine-scale model. We showed that by introducing an effective rigidity to the coarse-grained model and by controlling the volume fraction, we could achieve very high gain in numerical (CPU-time) efficiency while maintaining a quantitative approximation and minimizing the loss in spatial and temporal resolution of the model. Using our efficient polymer model one can simulate chromosome dynamics during the whole cell cycle on a desktop computer within a day. While we illustrated our approach using chromosomes on a lattice model as toy examples, our strategy is generic and can be applied to any self-avoiding polymers and to off-lattice systems. We emphasized that neglecting topological constraints may lead to an erroneous description of the fine-scale model. Therefore the effect of supplementary interactions added to the null model, to describe specific observations present in experimental data, may lead to misinterpretation.

In the second part, we used our strategy to build a coarse-grained null model for chromatin and decorated it with a copolymer framework based solely on epigenomic data to investigate the folding and dynamics of chromosome 3R of drosophila. It is the first study trying to quantitatively describe the behavior and time evolution of an entire chromosome (20 Mbp) during one cell cycle (20h of real time) with high precision (10 kbp resolution). Our heteropolymer model has one unique parameter, the strength of short-range interaction E_i between genomic loci having the same epigenomic state. Our findings are in qualitative agreement with our previous works based on shorter pieces of chromosomes[20,32], but significantly improve our description of chromosome folding in drosophila. By varying E_i, we showed that the system continuously switches from a dynamic homogeneous crumpled-like behavior to a crumpled heterogeneous micro-phased state. Interestingly, we observed that during this transition, the chromosome fluctuations characterized by the mean squared displacement conserve the same scaling behavior ($g_1(t) \approx \gamma t^{0.4}$) with exponents compatible with a crumpled polymer. However, the prefactor γ depends on E_i and is sharply reduced above a given strength of interaction, characteristics of a glass-like transition[66,67]. Another interesting observation was that the sequence-average contact probability $P_c(s)$, a quantity directly comparable to experimental data, is independent of E_i (at least for weak, biologically relevant values) and is same as in the reference null model, as already observed by Gursoy et al[69]. This motivates, afterward, the validity of homopolymer models, extensively used by polymer physicists, to study the generic physical principles behind chromosome folding based on comparison with sequence-average experimental data[16,39,41,57], even if such chromatin organization is strongly heterogeneous. This also suggests that, before adding specific interactions to the system, any quantitative approach should first aim to describe such sequence-average behaviors in a null homogeneous model.

Comparing our model predictions for chromosome 3R to the corresponding HiC data, we observed an excellent agreement at all scales, strongly suggesting that epigenomics is a primary driver of chromosome folding in drosophila. The strength of interaction compatible with the data is weak ($\sim 0.1\ kT$) and locate the *in vivo* situation in the transition zone between the

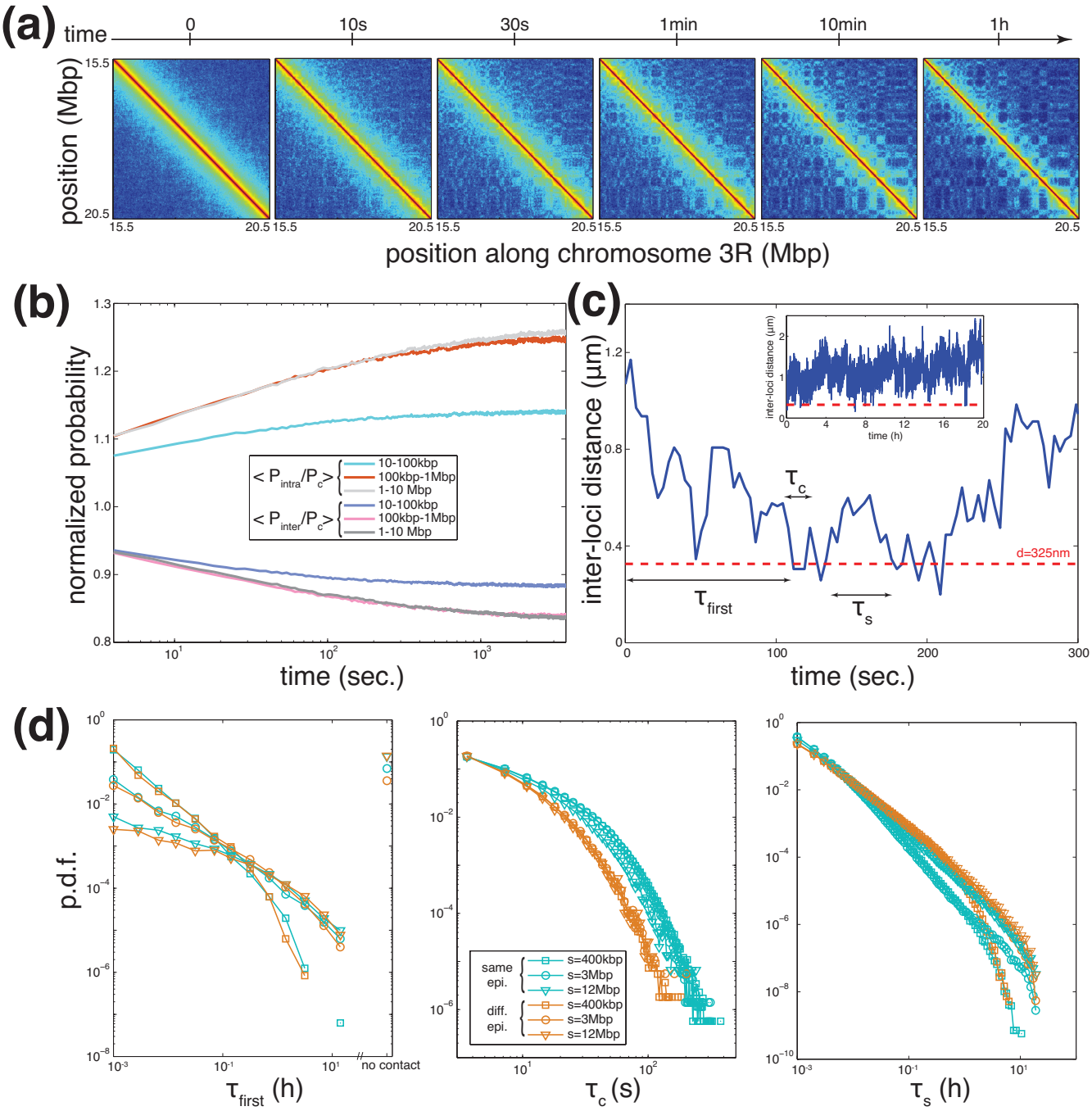

Figure 7. Dynamics of interactions. (a) Predicted contact maps ($E_i = -0.1\ kT$) for the region located between 15.5 and 20.5 Mbp of chromosome 3R as a function of time along the cell cycle. Same legend as in Fig.5c. (b) Time evolution of the ratio between P_{intra} and P_c (increasing curves) and of the ratio between P_{inter} and P_c (decreasing curves) averaged over genomic distances between 10 and 100 kbp, between 100 kbp and 1 Mbp and between 1 and 10 Mbp. (c) Example of the time evolution of distance between two loci in early times. The inset is the full trajectory along the cell cycle. The red dashed line represents the cut-off distance we choose to define that the two loci are in "contact" or not. From each trajectory, we define one value for the time of first encounter τ_{first} and several values for the contact time τ_c and the search time τ_s. (d) Probability distribution functions (p.d.f) of τ_{first} (left), τ_c (center) and τ_s (right) for pairs with the same (blue) or different (orange) epigenomic state separated by different genomic distance: $s = 400$ kbp (squares), 3 Mbp (circles) or 12 Mbp (triangles).

homogeneous crumpled and the micro-phased states. TADs are only partially collapsed and interact dynamically with other TADs of the same main epigenomic state. This suggests a substantial stochasticity in chromosome organization, consistent with recent single-cell HiC or super-resolution experiments[70–75]. We also detected several discrepancies between the predicted and experimental contact maps. For example, the model predicts spurious contacts or misses some between some TADs. This could be due to a wrong annotation of the local epigenomic state or the existence of specific interactions driven by other biological processes not accounted in the model. For example, refining the model to account more precisely for the local epigenetic content (for example the relative levels of histone modifications or chromatin-binding proteins) or differences in interaction strengths between different states would certainly lead to a better correspondence. We also observed that TADs are more sharply defined in the experiments, particularly in the corners of large TADs. This might be the results of an experimental bias inherent to the HiC technology or of the presence of cis-interacting mechanisms, like the recently proposed loop extrusion model in mammals[26,27], that enhance the contact frequency along the genome.

To exploit the capacity of the model to simulate long trajectories (20 hrs), we analyzed the time evolution of chromosome organization. TADs formed very quickly (within minutes), entirely consistent with HiC data made on synchronized mammalian cells showing that, in early G1, TADs are already observable in the data[74,76]. Formation of long-range contacts require more time and eventually appear hours after the mitotic exit, also consistent with the time evolution of HiC data during cell cycle in human[74,76] and yeast[77]. To go deeper into this characterization and get insights into the dynamics of contact formation, we tracked pairs of loci. At the investigated resolution (10 kbp), contacts are transient and their typical lifetime is ~ 10 seconds, indicating a very dynamic situation, consistent with many experiments performed on living cells[78]. Probability distribution functions of the first encounter time, the contact time, or the search time are polynomial, suggesting a possible connection with fractional brownian motion physics, as for bacterial chromosomes[79]. Interestingly, we predicted that a significant proportion ($5-15\%$) of long-range contacts (> 1 Mbp) are not established within one cell cycle. This suggests that the genomic distance between regulatory elements should not exceed 1 Mbp to ensure that physical contact between the elements, prerequisite to an activation or repression event, would happen at least once in the cell cycle in order to maintain a stable regulation or gene expression. With the recent progress in genome editing[80], it would be interesting to experimentally test such predictions by simultaneously tracking the distance between a promoter and its enhancer and the current gene activity[81], for various genomic distances between the two elements. All this suggests that the 3D chromosome organization in higher eucaryotes is out-of-equilibrium and the chromatin is very dynamic. This emphasizes the necessity to properly account for the time evolution of such organization in quantitative models of chromosomes, especially for higher eucaryotes where chromosomes are strongly topologically constrained.

As a proof of concept, we demonstrated the utility of our coarse-graining approach to study chromatin organization in drosophila. However, the numerical efficiency of the method opens new perspectives to investigate the physical and mechanistic principles behind chromosome folding more deeply with many aspects remain to be understood. For example, extrapolating to the whole human genome, it would require ~ 120h of CPU time with our strategy at 10kbp resolution to simulate one cell cycle while it remains illusive to do it with the fine-scale model (> 100 years of CPU time). The possibility to easily simulate the dynamics of chromosomes or full genomes during long biologically relevant time period would allow to quantitatively investigate in the future the role of other types of interactions, like those associated with the nuclear membrane, another important player in organizing chromosomes inside the nucleus[25,82] and the crosstalk with epigenomic-driven interaction as presented here. This situation seems particularly attractive to describe the reorganization of chromatin during senescence where constitutive heterochromatin detaches from the membrane to form large foci at the interior of the nucleus[83].

4 Materials and Methods

4.1 Simulation of structural and dynamical properties

The polymer is modeled as a semi-flexible self-avoiding walk, consisting of N beads, on a face centered cubic (fcc) lattice. Rigidity is accounted using a standard formulation:

$$\mathscr{H} = \frac{\kappa}{2}\sum_{i=1}^{N-1}(1-\cos\theta_i), \tag{4}$$

where κ is the bending rigidity and is directly related to $N_K = l_k/b$, and θ_i is the angle between the bond vectors i and $i+1$. Confinement and effect of other chains are approximated by using periodic boundary condition in a fcc lattice of size $S\times S\times S$, where each unit cell contains 4 lattice sites. The corresponding lattice volume fraction is $\Phi = N/(4S^3)$. Dynamics of the chain follows the local move scheme developed by Hugouvieux and coworkers[84]. This scheme allows two monomers to occupy the same lattice site if and only if they are consecutive along the chain. One Monte Carlo step (MCS) consists of N trial moves where a monomer is randomly chosen and randomly displaced to nearest-neighbor sites on the lattice (Fig.8). Trial moves are accepted according to a Metropolis criterion applied to $\mathscr{H}$ and if the chain connectivity is maintained. These simple

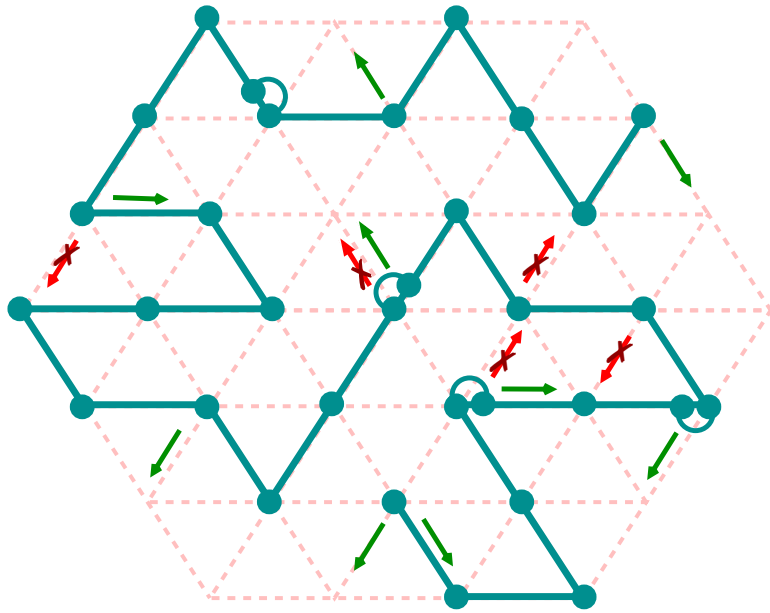

Figure 8. Schematic of the lattice polymer configuration on a 2D projection of a 3D fcc lattice. Solid line with beads represents the polymer, and the dotted lines represent the underlying lattice. Each lattice site is allowed to contain a maximum of two beads if they are consecutive to each other along the chain. Semicircular arcs indicate doubly occupied lattice sites. Some of the allowed and forbidden moves are shown in green and red respectively.

rules allow efficient simulations of reptation motion in dense - topologically constrained - systems, while still accounting for the main characteristics of polymer dynamics like polymer connectivity, excluded volume, and non-crossability of polymer strands. More details on the lattice model and KMC scheme can be found in[32,84].

In the entangled regime ($L/L_e \gg 1$), dynamics could be very slow and the system may keep the "memory" of its initial configuration and topology for a long time. Chromosomes are thought to be mostly knotted-free structures[16,41,85]. Therefore, we initiate our simulations by knotted-free configurations generated using the "hedgehog" algorithm[58,86,87]: starting from a central unknotted scaffold, configurations are iteratively grown by randomly inserting monomers at nearest-neighbor sites common to two already placed consecutive monomers (Fig.S1a).

Starting from a given initial configuration, we then normally simulate $10^7 - 10^8$ MCS and store the configurations after each 10^3 MCS. In some special cases where we are specifically interested in small time scales, we collect configurations more frequently. From these snapshots taken from 10^2 simulated trajectories, we then estimate several structural and dynamical quantities of interest. We focus on the time-evolution of the mean squared displacements (MSD) of individual monomers ($g_1(t)$) and of the center of mass of the chain ($g_3(t)$), as well as the average squared distance $\langle R^2(s)\rangle$ and contact probability $P_c(s)$ between monomers separated by a linear distance s along the chain. For the latter, a contact is defined if the physical distance between a pair of genomic loci is less than a particular distance d_c.

4.2 Relation between Kuhn length l_K and lattice parameters

In this section we derived the relation between Kuhn length l_K and different lattice parameters, expressed in Eq.3. We consider an arbitrary fine-scale model (FSM) of a semi-flexible self-avoiding walk defined by N_0, l_k^0, b_0. We note N, l_k and b, the corresponding values of a coarse-grained model (CGM) of the FSM. Each CGM monomer encompasses $n = N_0/N > 1$ FSM monomers.

Using Eq.1, conservation of L/L_e and conservation of the volumic density in FSM monomer ρ leads to

$$\frac{L}{L_e} = \frac{Nb}{l_K}\left(\frac{\rho b\, l_k^2}{c}\right)^2 \tag{5}$$

$$\equiv \frac{L_0}{L_e^0} = \frac{N_0 b_0}{L_e^0} \tag{6}$$

$$\left(\frac{l_K}{b}\right)^3 = \left(\frac{nc^2}{\rho^2 b^6}\right)\left(\frac{b_0}{L_e^0}\right) \tag{7}$$

$\rho = (N_0/V)$ with V the volume of the box. Conservation of the volume implies that

$$\rho = \frac{N\, n\sqrt{2}}{N_s b^3} = \frac{\sqrt{2}\Phi\, n}{b^3} \tag{8}$$

with N_s the number of lattice sites and $\Phi = N/N_s$ the lattice volumic fraction. Incorporating Eq.8 into Eq.7 leads to

$$\frac{l_k}{b} = \left[\left(\frac{b_0}{L_e^0}\right)\left(\frac{c^2}{2n\Phi^2}\right)\right]^{1/3} \tag{9}$$

Practically, knowing ρ, b_0 and L_e^0 (from Eq. 1 of the main text) for the reference model, for a given coarse-graining (defined by n), Eq.9 gives us a relation between l_k/b (which is related to the bending energy of the model, see below) and the lattice volumic fraction Φ. Eq.8 is used to compute the corresponding value for b.

4.3 Relation between Kuhn size N_k and bending rigidity κ for lattice polymer

For a lattice phantom chain with N beads, the mean squared end-to-end distance $\langle R_e^2 \rangle$ is given by

$$\langle R_e^2 \rangle = \frac{12}{13} b^2 \left[(N-1)\left(\frac{1+x}{1-x}\right) - \frac{2x\left(1-x^{N-1}\right)}{(1-x)^2} \right] \tag{10}$$

where

$$x = \frac{12}{13}\langle \cos\theta \rangle \tag{11}$$

$$= \frac{12}{13}\left(\frac{1+2exp[-\kappa/2]-2exp[-3\kappa/2]-exp[-2\kappa]}{1+4\exp[-\kappa/2]+2exp[-k]+4\exp[-3\kappa/2]+exp[-2\kappa]}\right), \tag{12}$$

where θ is the angle between two monomers and κ is the bending rigidity correspond to the bending energy $E(\theta) = \kappa(1-\cos\theta)$. Now we have $l_K/b = (1+x)/(1-x)$, which relates N_K to κ.

Acknowledgements

We thank Magali Richard, Ivan Junier and Cédric Vaillant for critical reading of the manuscript, as well as Ralf Everaers, Pascal Carrivain, Aurélien Bancaud, Mikhail Tamm and Giacomo Cavalli for fruitful discussions. We acknowledge the CIMENT infrastructure (supported by the Rhône-Alpes region, Grant CPER07_13 CIRA) for computing resources. This work was supported by Agence Nationale de la Recherche (ANR-15-CE12-0006 EpiDevoMath), Fondation pour la Recherche Médicale (DEI20151234396) and CNRS.

References

1. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015). URL http://dx.doi.org/10.1016/j.cell.2015.04.004. DOI 10.1016/j.cell.2015.04.004.
2. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nat.* **538**, 265–269 (2016).
3. Jost, D., Vaillant, C. & Meister, P. Coupling 1d modifications and 3d nuclear organization: data, models and function. *Curr Opin Cell Biol* **44**, 20–27 (2017).
4. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001). URL http://dx.doi.org/10.1038/35066075. DOI 10.1038/35066075.
5. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Sci.* **326**, 289–293 (2009). URL http://dx.doi.org/10.1126/science.1181369. DOI 10.1126/science.1181369.
6. Rao, S. S. P. *et al.* A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–80 (2014). DOI 10.1016/j.cell.2014.11.021.
7. Ho, J. W. K. *et al.* Comparative analysis of metazoan chromatin organization. *Nat.* **512**, 449–52 (2014). DOI 10.1038/nature13415.

8. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nat.* **485**, 376–380 (2012).

9. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nat.* **485**, 381–385 (2012).

10. Zhan, Y. *et al.* Reciprocal insulation analysis of hi-c data shows that tads represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Res.* **27**, 479–490 (2017).

11. Ulianov, S. V. *et al.* Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70–84 (2016).

12. Zhu, Y. *et al.* Constructing 3d interaction maps from 1d epigenomes. *Nat. Commun.* **7**, 10812 (2016).

13. Mourad, R. & Cuvier, O. Computational identification of genomic features that influence 3d chromatin domain formation. *PLoS Comput. Biol.* **12**, e1004908 (2016).

14. Haddad, N., Vaillant, C. & Jost, D. Ic-finder: inferring robustly the hierarchical organization of chromatin folding. *Nucleic Acids Res* **45**, e81 (2017).

15. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–59 (2015). DOI 10.1016/j.cell.2015.02.040.

16. Rosa, A. & Everaers, R. Structure and dynamics of interphase chromosomes. *PLOS Comput. Biol.* **4**, 1–10 (2008).

17. Hyeon, C. & Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2** (2011).

18. Barbieri, M. *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A* **109**, 16173–16178 (2012). URL http://dx.doi.org/10.1073/pnas.1204799109. DOI 10.1073/pnas.1204799109.

19. Sharma, R. *et al.* Differential spatial and structural organization of the x chromosome underlies dosage compensation in c. elegans. *Genes Dev* **28**, 2591–2596 (2014). URL http://dx.doi.org/10.1101/gad.248864.114. DOI 10.1101/gad.248864.114.

20. Jost, D., Carrivain, P., Cavalli, G. & Vaillant, C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* **42**, 9553 (2014).

21. Tark-Dame, M., Jerabek, H., Manders, E. M. M., Heermann, D. W. & van Driel, R. Depletion of the chromatin looping proteins ctcf and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. *PLoS Comput. Biol* **10**, e1003877 (2014). DOI 10.1371/journal.pcbi.1003877.

22. Ganai, N., Sengupta, S. & Menon, G. I. Chromosome positioning from activity-based segregation. *Nucleic Acids Res.* **42**, 4145–4159 (2014).

23. Benedetti, F., Dorier, J., Burnier, Y. & Stasiak, A. Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. *Nucleic Acids Res.* **42**, 2848–2855 (2014).

24. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).

25. Jerabek, H. & Heermann, D. W. How chromatin looping and nuclear envelope attachment affect genome organization in eukaryotic cell nuclei. *Int Rev Cell Mol Biol* **307**, 351–381 (2014). URL http://dx.doi.org/10.1016/B978-0-12-800046-5.00010-2. DOI 10.1016/B978-0-12-800046-5.00010-2.

26. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456–E6465 (2015). URL http://dx.doi.org/10.1073/pnas.1518552112. DOI 10.1073/pnas.1518552112.

27. Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell Rep* **15**, 2038–2049 (2016). URL http://dx.doi.org/10.1016/j.celrep.2016.04.085. DOI 10.1016/j.celrep.2016.04.085.

28. Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. Compaction and segregation of sister chromatids via active loop extrusion. *Elife* **5** (2016).

29. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nat.* **529**, 418–422 (2016).

30. Pierro, M. D., Zhang, B., Aiden, E. L., Wolynes, P. G. & Onuchic, J. N. Transferable model for chromosome architecture. *Proc Natl Acad Sci U S A* **113**, 12168–12173 (2016).

31. Wani, A. H. *et al.* Chromatin topology is coupled to polycomb group protein subnuclear organization. *Nat Commun* **7**, 10291 (2016). URL http://dx.doi.org/10.1038/ncomms10291. DOI 10.1038/ncomms10291.

32. Olarte-Plata, J. D., Haddad, N., Vaillant, C. & Jost, D. The folding landscape of the epigenome. *Phys. Biol.* **13**, 026001 (2016).

33. Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer physics of chromosome large-scale 3d organisation. *Sci. Rep.* **6**, 29775 (2016).

34. Brackley, C. A., Johnson, J., Kelly, S., Cook, P. R. & Marenduzzo, D. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* **44**, 3503–3512 (2016).

35. Brackley, C. A. *et al.* Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol.* **17**, 59 (2016).

36. Tiana, G. *et al.* Structural fluctuations of the chromatin fiber within topologically associating domains. *Biophys. J.* **110**, 1234–1245 (2016).

37. Florescu, A.-M., Therizols, P. & Rosa, A. Large scale chromosome folding is stable against local changes in chromatin structure. *Plos. Comput. Biol.* **12**, e1004987 (2016).

38. Hyeon, C. & Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2**, 487 (2011).

39. Halverson, J. D., Smrek, J., Kremer, K. & Grosberg, A. Y. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Rep. Prog. Phys.* **77**, 022601 (2014).

40. Cremer, T. *et al.* Chromosome territory organization within the nucleus. *Rev. Cell Biol. Mol. Medicine* (2012).

41. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosom. Res.* **19**, 37–51 (2011).

42. Putz, M., Kremer, K. & Grest, G. What is the entanglement length in a polymer melt? *Eur. Lett.* **49**, 735 (2000).

43. Uchida, N., Grest, G. S. & Everaers, R. Viscoelasticity and primitive path analysis of entangled polymer liquids: From f-actin to polyethylene. *The J. Chem. Phys.* **128**, 044902 (2008).

44. De Gennes, P. G. Scaling concepts in polymer physics. *Cornell Univ. Press.* **32**, 290–290 (1980).

45. Hajjoul, H. *et al.* High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res.* **23(11)**, 1829–1838 (2013).

46. Münkel, C. & Langowski, J. Chromosome structure predicted by a polymer model. *Phys. Rev. E* **57**, 5888–5896 (1998). DOI 10.1103/PhysRevE.57.5888.

47. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. Bionumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38**, D750–3 (2010). DOI 10.1093/nar/gkp889.

48. Cheutin, T. & Cavalli, G. Progressive polycomb assembly on h3k27me3 compartments generates polycomb bodies with developmentally regulated motion. *Plos Genet.* **8**, e1002465 (2012).

49. Doi, M. & Edwards, S. F. The theory of polymer dynamics. *Clarendon Press.* **32**, 290–290 (1988).

50. Tamm, M. V., Nazarov, L. I., Gavrilov, A. A. & Chertovich, A. V. Anomalous diffusion in fractal globules. *Phys. Rev. Lett.* **114**, 178102 (2015).

51. Farge, E. & Maggs, A. C. Dynamic scattering from semiflexible polymers. *Macromol.* **26**, 5041–5044 (1993).

52. Bystricky, K., Heun, P., Gehlen, L., Langowski, J. & Gasser, S. M. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. USA* **101**, 16495–16500 (2004).

53. Kimura, H. *et al.* The genome folding mechanism in yeast. *J. Biochem.* **154**, 137–147 (2013).

54. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nat.* **465**, 363–367 (2010).

55. Grosberg, A., Rabin, Y., Havlin, S. & Neer, A. Crumpled globule model of the three-dimensional structure of DNA. *Eur. Lett.* **23**, 373–378 (1993).

56. Rosa, A., Becker, N. B. & Everaers, R. Looping probabilities in model interphase chromosomes. *Biophys. J.* **98**, 2410–2419 (2010).

57. Rosa, A. & Everaers, R. Ring polymers in the melt state: the physics of crumpling. *Phys. Rev. Lett.* **112**, 118302 (2014).

58. Imakaev, M. V., Tchourine, K. M., Nechaev, S. K. & Mirny, L. A. Effects of topological constraints on globular polymers. *Soft Matter* **11**, 665–671 (2015).

59. Lucas, J. S., Zhang, Y., Dudko, O. K. & Murre, C. 3d trajectories adopted by coding and regulatory dna elements: First-passage times for genomic interactions. *Cell* **158**, 339–352 (2014).

60. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the drosophila genome. *Cell* **148**, 458 – 472 (2012).

61. Haddad, N., Jost, D. & Vaillant, C. Perspectives: using polymer modeling to understand the formation and function of nuclear compartments. *Chromosom. Res.* **25**, 35–50 (2017). URL http://dx.doi.org/10.1007/s10577-016-9548-2. DOI 10.1007/s10577-016-9548-2.

62. Canzio, D. *et al.* A conformational switch in hp1 releases auto-inhibition to drive heterochromatin assembly. *Nat.* **496**, 377–381 (2013). URL http://dx.doi.org/10.1038/nature12032. DOI 10.1038/nature12032.

63. Isono, K. *et al.* Sam domain polymerization links subnuclear clustering of prc1 to gene silencing. *Dev Cell* **26**, 565–577 (2013). URL http://dx.doi.org/10.1016/j.devcel.2013.08.016. DOI 10.1016/j.devcel.2013.08.016.

64. Larson, A. G. *et al.* Liquid droplet formation by hp1 suggests a role for phase separation in heterochromatin. *Nat.* (2017).

65. Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in drosophila cells. *Cell* **143**, 212 – 224 (2010).

66. Donth, E.-J. *The Glass Transition: Relaxation Dynamics in Liquids and Disordered Materials* (Springer-Verlag, 2001).

67. Mezard, M., Parisi, G. & Virasoro, M. *Spin glass theory and beyond* (World Scientific, Singapore, 1987).

68. Miles, I. S. & Rostami, S. (eds.) *Multicomponent Polymer Systems* (Longman Scientific and Technica, Singapore, 1992).

69. Gürsoy, G., Xu, Y., Kenter, A. L. & Liang, J. Spatial confinement is a major determinant of the folding landscape of human chromosomes. *Nucleic Acids Res.* **42**, 8223–8230 (2014).

70. Nagano, T. *et al.* Single-cell hi-c reveals cell-to-cell variability in chromosome structure. *Nat.* **502**, 59–64 (2013). URL http://dx.doi.org/10.1038/nature12593. DOI 10.1038/nature12593.

71. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Sci.* **353**, 598–602 (2016).

72. Stevens, T. J. *et al.* 3d structures of individual mammalian genomes studied by single-cell hi-c. *Nat.* **544**, 59–64 (2017).

73. Flyamer, I. M. *et al.* Single-nucleus hi-c reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nat.* **544**, 110–114 (2017).

74. Nagano, T. *et al.* Cell cycle dynamics of chromosomal organisation at single-cell resolution. *bioRxiv* https://doi.org/10.1101/094466 (2017).

75. Cattoni, D. I. *et al.* Single-cell absolute contact probability detection reveals that chromosomes are organized by modulated stochasticity. *bioRxiv* https://doi.org/10.1101/159814 (2017).

76. Naumova, N. *et al.* Organization of the mitotic chromosome. *Sci.* **342**, 948–953 (2013). DOI 10.1126/science.1236083.

77. Lazar-Stefanita, L. *et al.* Choreography of budding yeast chromosomes during the cell cycle. *bioRxiv* https://doi.org/10.1101/096826 (2017).

78. Bystricky, K. Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. *FEBS Lett.* **589**, 3014–3022 (2015).

79. Polovnikov, K., Gherardi, M., Cosentino-Lagomarsino, M. & Tamm, M. Folding and cytoplasm viscoelasticity contribute jointly to chromosome dynamics. *arXiv* arXiv:1703.10841 (2017).

80. Sander, J. D. & Joung, J. K. Crispr-cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).

81. Chen, H., Fujioka, M. & Gregor, T. Direct visualization of transcriptional activation by physical enhancer-promoter proximity. *bioRxiv* http://dx.doi.org/10.1101/099523 (2017).

82. Mattout, A., Cabianca, D. S. & Gasser, S. M. Chromatin states and nuclear organization in development— a view from the nuclear lamina. *Genome Biol.* **16**, 174 (2015).

83. Chandra, T. *et al.* Global reorganization of the nuclear landscape in senescent cells. *Cell Rep* **10**, 471–483 (2015).

84. Hugouvieux, V., Axelos, M. & Kolb, M. Amphiphilic multiblock copolymers: From intramolecular pearl necklace to layered structures. *Macromol.* **42**, 392–400 (2009).

85. Zhang, B. & Wolynes, P. G. Topology, structures, and energy landscapes of human chromosomes. *Proc Natl Acad Sci U S A* **112**, 6062–6067 (2015).

86. Earnshaw, W. C. & Laemmli, U. K. Architecture of metaphase chromosomes and chromosome scaffolds. *The J. Cell Biol.* **96**, 84–93 (1983). DOI 10.1083/jcb.96.1.84.

87. Klenin, K. V., Vologodskii, A. V., Anshelevich, V. V., Dykhne, A. M. & Frank-Kamenetskii, M. D. Effect of excluded volume on topological properties of circular dna. *J. Biomol. Struct. Dyn.* **5**, 1173–1185 (1988).