



**HAL**  
open science

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

Surya Ghosh, Daniel Jost

► **To cite this version:**

Surya Ghosh, Daniel Jost. How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes. PLoS Computational Biology, 2018, 14 (5), pp.e1006159. 10.1371/journal.pcbi.1006159 . hal-01976607

**HAL Id: hal-01976607**

**<https://hal.science/hal-01976607v1>**

Submitted on 13 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

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

Surya K. Ghosh\*, Daniel Jost\*

Univ Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, F-38000 Grenoble, France

\* [surya.ghosh@univ-grenoble-alpes.fr](mailto:surya.ghosh@univ-grenoble-alpes.fr) (SKG); [daniel.jost@univ-grenoble-alpes.fr](mailto:daniel.jost@univ-grenoble-alpes.fr) (DJ)



**OPEN ACCESS**

**Citation:** Ghosh SK, Jost D (2018) How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes. *PLoS Comput Biol* 14(5): e1006159. <https://doi.org/10.1371/journal.pcbi.1006159>

**Editor:** Sheng Zhong, University of California, San Diego, UNITED STATES

**Received:** October 9, 2017

**Accepted:** April 28, 2018

**Published:** May 29, 2018

**Copyright:** © 2018 Ghosh, Jost. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research is supported by Agence Nationale de la Recherche (ANR-15-CE12-0006 EpiDevoMath), Fondation pour la Recherche Medicale (DEI20151234396), Centre national de la recherche scientifique (CNRS). We acknowledge computational resources from CIMENT infrastructure (supported by the Rhone-Alpes region, Grant CPER07 13 CIR). The funders had

## Abstract

The 3D organization of chromosomes 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 challenges are therefore to develop adequate models and simulation methods to investigate properly the multi spatio-temporal scales of such macromolecules. Here, we proposed a generic strategy to develop efficient coarse-grained models for self-avoiding polymers on a lattice. 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.

## Author summary

The chromosome architecture inside cell nuclei plays important roles in regulating cell functions. Many experimental and modeling efforts are dedicated to deciphering the mechanisms controlling such organization. There are proliferations of experimental studies which report the hierarchical structure of chromosomes but how exactly they physically organize in 3D is not fully understood. In modeling, the main challenges are to develop adequate models and simulation methods to investigate correctly these highly dense long polymer chains. Taken into consideration the fundamental physical characteristics of chromosomes, we developed robust and numerically efficient polymer models

no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

that enabled us to explore the dynamics of long chromosomes over long time periods with good statistics. We applied this framework to investigate the dynamical folding of chromosome in drosophila. Accounting for the local biochemical information, we were able to reproduce the experimentally-measured contact frequencies between any pairs of genomic loci quantitatively and to track the hierarchical chromosome structure throughout the cell cycle. Our results further support the picture of a very dynamic chromosome organization driven by weak short-range interactions.

## Introduction

Though all cells of a multicellular organism contain the same genetic information, they vary widely in shapes, physiologies, and 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 its 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 (Hi-C) 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 section ‘Application to chromatin folding in drosophila’ for 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 understand 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, 16–39]. An important challenge for such models is to simulate, with 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 [40]. 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 essential role in controlling the

dynamics and fluctuations of such polymers [16, 41]. However, when building their null models, very few approaches account adequately of such constrained situations. 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 develop coarse-grained null models for chromosomes. In the second part, we apply it to investigate the role of epigenomic-driven interactions in the folding and dynamics of drosophila genome. Finally, we discuss our results and their implications in the general context of chromosome modeling and chromatin biology.

## Results

### Chromosome and entanglement length

Chromosomes are long polymers confined inside a small volume, the cell nucleus [42]. As a result, the generic characteristics of these densely packed long polymers are very different from free isolated chains and exhibit distinct universal properties [41, 43]. For a simple semi-flexible self-avoiding polymer, composed of  $N$  beads of size  $b$  (in  $nm$ ), each bead representing  $\nu bp$ , such properties are mainly determined by (i) its rigidity, characterized by its Kuhn length  $l_k$  (in  $nm$ ), and (ii) by its volumic density  $\rho$  (in  $bp/nm^3$ ). 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 the polymer contour length  $L \equiv Nb$  and  $L_e$  (in  $nm$ ), the so-called entanglement length.  $L_e$  measures the typical sub-chain size above which topological confinement due to excluded volume influences configurational fluctuations, and depends on  $l_k$  and  $\rho$ . 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 [44], and may be estimated using the phenomenological relation [45]

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

with  $c \approx 19$  a numerical factor and  $\rho_k = (\rho/\nu)(b/l_k)$  the volumic density in Kuhn segment. In the following, we define  $N_k = \nu l_k/b$  (in  $bp$ ) as the Kuhn segment size representing the number of  $bp$  in one Kuhn segment.  $L/L_e \ll 1$  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, the equilibration time of the chain scales as  $N^3$  [46], 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.

**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  $l_k \equiv l_k^0$ , number of beads  $N \equiv N_0$  and bead size  $b \equiv b_0$  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 [47], the so-called 10 *nm*-fiber, to hundreds of nanometers, the so-called 30 *nm*-fiber [48]. We decide to use, as a reference model, the nucleosomal scale (1 monomer correspond to  $\nu_0 = 200$

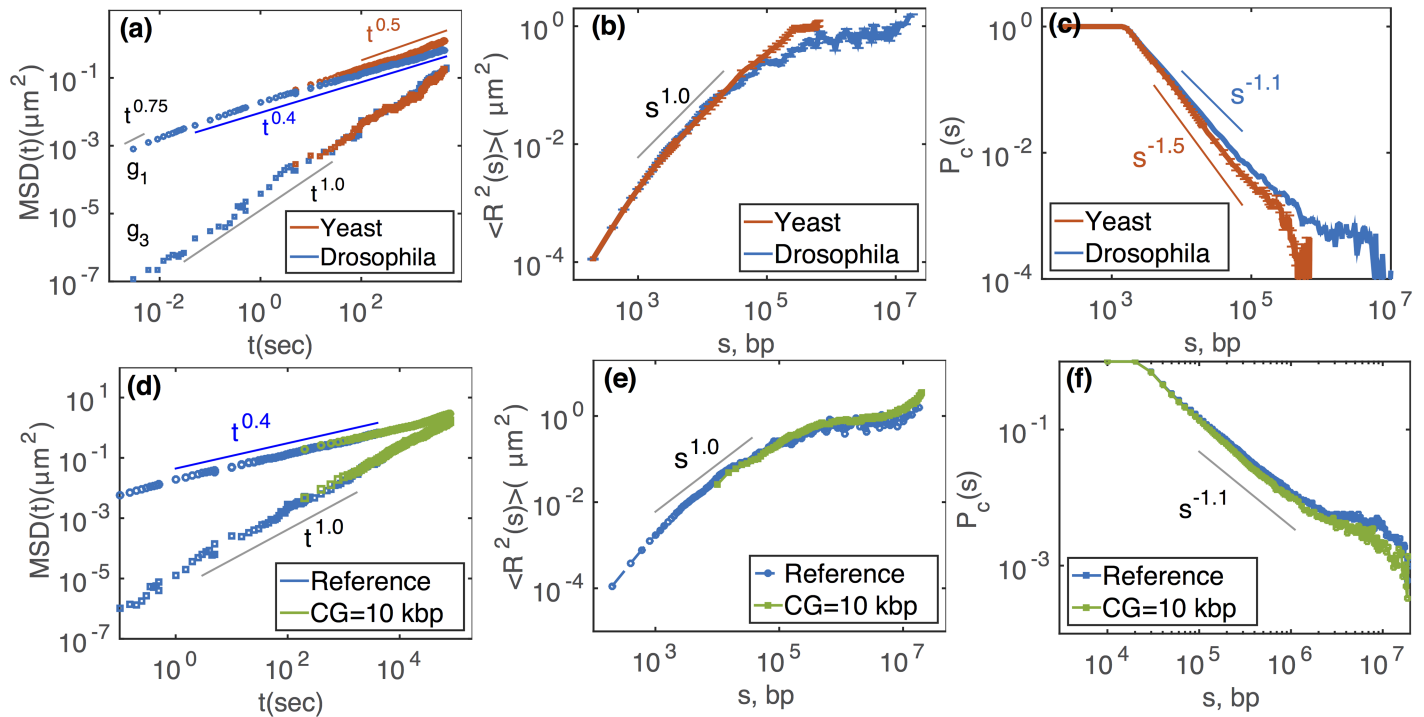
bp,  $b_0 \approx 10 \text{ nm}$ ) with a recent estimation of Kuhn size  $N_k^0 = v_0 l_k^0 / b_0 \approx 1 \text{ kbp}$  ( $l_k^0 / b_0 = 5$  monomers) based on cyclization probabilities of chromatin [26].

In order to determine the value of the entanglement length  $L_e$ , the other important quantity to fix is the chromatin volumic density  $\rho$  defined as the ratio between the genome size and the volume of the nucleus. Depending on the species, the cell types or the developmental stages, it may strongly vary. Typical orders of magnitude are  $\rho \approx 0.005 \text{ bp/nm}^3$  for haploid yeast,  $\rho \approx 0.009 \text{ bp/nm}^3$  for drosophila late embryos or  $\rho \approx 0.015 \text{ bp/nm}^3$  in typical mammalian nuclei (see [Materials and methods](#)). Systems with higher volumic density become more entangled and exhibit shorter entanglement length (Eq 1). This leads to decreasing values for  $N_e \equiv v_0 L_e / b_0 \approx 920 \text{ kbp}$  (yeast),  $285 \text{ kbp}$  (drosophila) or  $102 \text{ kbp}$  (mammals). In yeast,  $v_0 N_0 \approx 750 \text{ kbp}$ , implying that chromosomes are weakly entangled ( $L/L_e = 0.8 \lesssim 1$ ). In higher eukaryotes, as drosophila or mammals, chromosomes are longer (tens or hundreds of Mbp) and in the regime of strong topological constraints ( $L/L_e \gg 1$ ). For example, for a chromosome of length  $v_0 N_0 = 20 \text{ Mbp}$ , the corresponding value is  $L/L_e = 70$  in drosophila. For a given species, the exact value of this ratio may vary depending on the cell types due to variation in the volume of the nucleus but usually the entanglement regime is preserved (weakly constrained for yeast, strongly for higher eukaryotes). Note that the  $30 \text{ nm}$  fiber model for chromatin ( $l_k^0 = 300 \text{ nm}$ ,  $b_0 = 30 \text{ nm}$ ,  $v_0 l_k^0 / b_0 = 30 \text{ kbp}$ ) would lead to similar orders of magnitude for  $L/L_e$ .

**Generic behavior of chromosome.** To illustrate the generic behavior of the reference model in the different entanglement regimes, we perform kinetic Monte-Carlo (KMC) simulations of the chain dynamics using a lattice model with periodic boundary conditions and starting from knot-free initial configurations (see [Materials and methods](#)). We focus on the “yeast” ( $N_0 = 3750$  monomers,  $v_0 N_0 = 750 \text{ kbp}$ , lattice density  $\Phi_0 \equiv \rho b_0^3 / (\sqrt{2}v) = 0.023$ ) and on the “drosophila” ( $N_0 = 10^5$  monomers,  $v_0 N_0 = 20 \text{ Mbp}$ ,  $\Phi_0 = 0.043$ ) cases.

During the simulations, we measured four standard physical quantities: time evolution of the average mean squared displacements (MSD) of individual monomers  $g_1(t)$ , MSD of the center of mass of the chain  $g_3(t)$ , average mean squared distance  $\langle R^2(s) \rangle$  between two monomers separated by a contour length  $s$  (in  $bp$ ) along the chain and contact probability  $P_c(s)$  (Fig 1). Comparison of  $g_1(t)$  with  $0.01 t^{0.5}$ , the experimentally measured typical value of  $g_1$  in  $\mu\text{m}^2$  for yeast and drosophila [47, 49], 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  $\mu\text{m}$ . For the drosophila case, a  $10^8$ -MCS long trajectory would correspond to about  $30 \text{ min}$  of real time. To check the precision of the MSD scaling laws, we calculated  $g_1, g_3$  by varying the measuring simulation time window ( $\Delta t$ ) of the trajectory. We observed that both  $g_1, g_3$  reached steady-state rapidly and almost perfectly overlap for different trace-lengths  $\Delta t$ , see Fig. B in [S1 Text](#).

The yeast chromosome behaves dynamically as a standard Rouse chain [47]. At intermediate times,  $g_1 \sim t^{0.5}$  (Fig 1a), the typical scaling law in the Rouse diffusion regime [50]. At the 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 [50]. 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 and 1d). 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 [46]. This is a characteristic of knot-free polymers, like crumpled or ring polymers [51] and is reminiscent of the initial unknotted configurations. Starting from random configurations that contain complex knots (Fig.A(g) in [S1 Text](#)), we recover the standard reptation regime (Fig.D(a,d) in [S1 Text](#)).



**Fig 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 the drosophila case. (a,d) Individual MSD  $g_1(t)$  (top curves), and center of mass MSD  $g_3(t)$  (bottom curves) as a function of time  $t$ . (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  $P_c(s)$  as a function of  $s$ . A contact between any two monomers is defined if the 3D distance is less than a threshold  $d_c$  (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. For the yeast case, we fix  $L/L_e = 0.8$ ,  $\rho = 0.005 \text{ bp/nm}^3$ , and for drosophila,  $L/L_e = 70$ ,  $\rho = 0.009 \text{ bp/nm}^3$  (see section ‘A reference model for chromosome’).

<https://doi.org/10.1371/journal.pcbi.1006159.g001>

At small time scales ( $t < \text{ms}$ ),  $g_1$  scales as  $t^{0.75}$  which corresponds to the typical diffusion regime of a semi-flexible chain up to the Kuhn length scale [52].

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, 43, 53–55]. The yeast case is fully consistent with a worm-like-chain at equilibrium with  $\langle R^2(s) \rangle \sim s^1$  and  $P_c \sim s^{-1.5}$  [50]. 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, 43, 56–58]. The large scale behavior ( $s > 1$  Mbp) is a remaining signature of the initial scaling laws (Fig. G in S1 Text): the system has yet to reach steady-state and is still strongly out-of-equilibrium.

### 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, 43, 51, 56, 57, 59] and with experiments [5, 6, 47, 49, 53–55, 60]. At this nucleosomal scale, the model has very good spatial (10 nm) and temporal (15  $\mu\text{sec}$ , 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.

**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  $N_k^0 \equiv n l_k^0 / b_0$ , 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  $b = \sqrt{n b_0 l_k^0}$ . 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  $\Phi$  occupied by the CG chain (for a lattice model  $\Phi \equiv N/N_s$  with  $N_s$  the number of lattice sites). 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  $\Phi_0$  is already high in the FSM and/or the coarse-graining is strong ( $n \gg 1$ ),  $\Phi$  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 choose  $b < \sqrt{n b_0 l_k^0}$  in order to maintain  $\Phi < 1$ , violating the conservation of the ratio  $L/L_e$ . This may affect the dynamical regime of the chain, and hence its physical properties (see section ‘How to build a good coarse-grained null model of chromosome’). 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  $l_k$ . And we imposed 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](#))

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

with  $\rho_{FS} \equiv N_0/V = \rho/v_0$  the volumic density in FS monomers (with  $V$  the volume of the box).  $\Phi$  now plays the role of a control parameter: the characteristics of the CGM depends not only on the FS properties but also on  $\Phi$  (see [Table 1](#) for examples) since a given  $\Phi$  determines  $b$  and  $l_k$ , and the corresponding value for the lattice bending energy  $\kappa$  is inferred from  $l_k/b$  (see [Materials and methods](#)). As in most coarse-graining approaches, the size of each CG monomer ( $b$ ) does not reflect the actual contour length of the corresponding fine-scale subchain, but rather would represent the typical diameter of the volume occupied by the fine-scale monomers. However, we observe that the length deformation of the CG polymer with respect to the reference model remains weak (Fig.P(a) in [S1 Text](#)).

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  $\Phi$ . Therefore, the CGM cannot pretend to quantitatively describe the FSM properties at scales smaller than few  $l_k$ .

**Table 1. Simulation parameters at different coarse-grainings for a semi-flexible, self-avoiding polymer with  $L/L_e = 70$  and  $\rho = 0.009 \text{ bp/nm}^3$  (drosophila case).** Lattice volumic fraction  $\Phi$ , Kuhn size  $N_k \equiv vl_k/b$  (in kbp), bond length of the polymer  $b$  (in nm) and Kuhn length  $l_k$  (in nm) at different coarse-grainings (CG) of  $\nu = 0.2, 2, 5$ , and 10 kbp. Each fifth subcolumn represents the time in msec equivalent to one Monte-Carlo steps (MCS). Similar tables for the yeast and mammalian cases are given in the supplementary text.

200 bp					2 kbp					5 kbp					10 kbp				
$\Phi_0$	$N_k^0$	$b_0$	$l_k^0$	MCS	$\Phi$	$N_k$	$b$	$l_k$	MCS	$\Phi$	$N_k$	$b$	$l_k$	MCS	$\Phi$	$N_k$	$b$	$l_k$	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					

<https://doi.org/10.1371/journal.pcbi.1006159.t001>

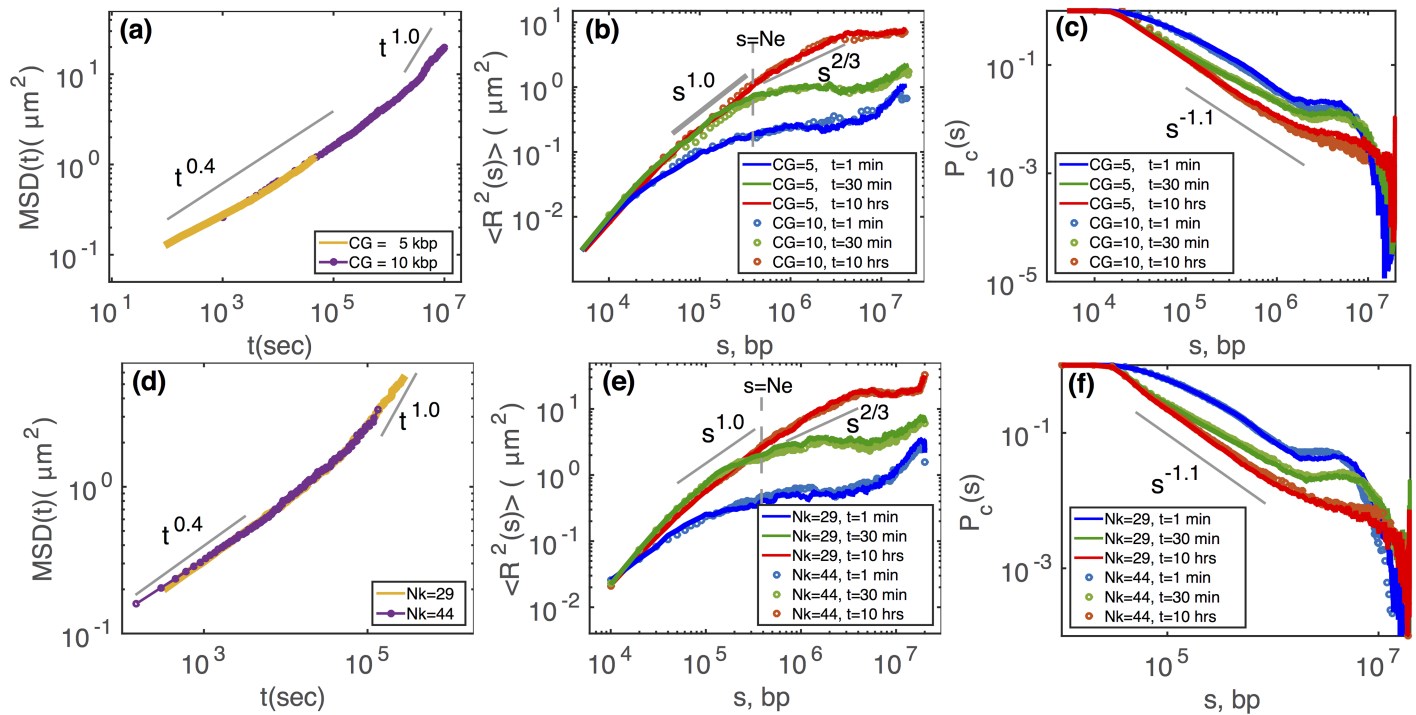
**Conservation of generic properties and time mapping.** In this part, we test if the above coarse-graining strategy conserves the structural and dynamical properties of the reference fine-scale model. In this article, we primarily focus on the drosophila case. However, in the Supplementary Information, we show that the method also performs very well for the yeast and mammalian cases (Fig.H, I in S1 Text) and that the success of the strategy does not depend on the type of used initial conditions (Fig.C, E in S1 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 section ‘Generic behaviour of chromosome’), 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 and 1f). For  $s < 1 \text{ Mbp}$ , we observe a perfect match between CG predictions and the FSM behavior. For  $s > 1 \text{ Mb}$ , the system does not reach steady-state and keeps a partial memory of the initial scaling laws that are different for both systems (Fig.F, G in S1 Text), 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  $N_K \equiv vl_k/b$  ( $N_K = 23 \text{ kbp}$ ): (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 and 2c). 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  $P_c(s)$  and the average second moment  $\langle \sigma^2(s) \rangle$  of the squared distance  $R^2(s)$  defined as  $\sigma^2(s) = \langle R^4(s) \rangle - \langle R^2(s) \rangle^2$  (Fig.J(c,d) in S1 Text).

To test that controlling the volumic fraction  $\Phi$ , 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 and Fig.J(a,b) in S1 Text). Identically, we observe almost perfect matching for the time-evolution of  $\langle R^2 \rangle$ ,  $P_c$  and  $\sigma^2(s)$  for all ranges of genomic distance. To push our strategy to the limit, we also considered very high values for  $\Phi$  (Fig.K(a,b,c) in S1 Text). In our lattice polymer model, each point can be occupied by two consecutive monomers (see section ‘Simulation of structural and dynamical properties of the null model’), so in principle, a maximum volumic fraction  $\Phi = 2$  is achievable. For  $\Phi \lesssim 1$ , all the





**Fig 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) for the drosophila case ( $L/L_e = 70, \rho = 0.009\text{bp}/\text{nm}^3$ ).** (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).

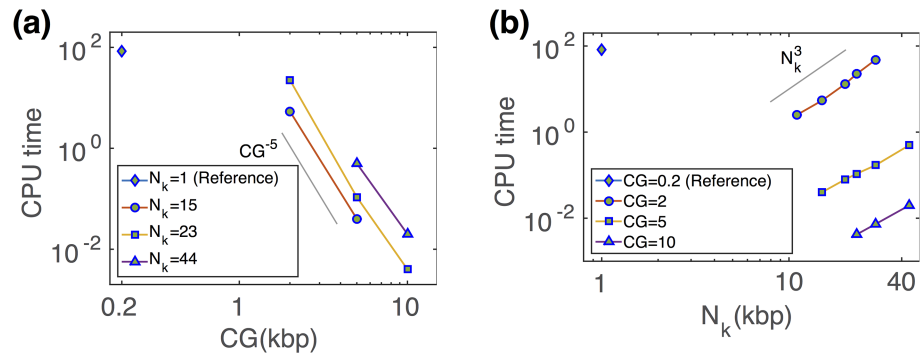
<https://doi.org/10.1371/journal.pcbi.1006159.g002>

simulations show exactly the same results as the reference model and follow the same curve. For  $\Phi \geq 1$  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  $\Phi$  (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  $\Phi$  which may impose restrictions on the dynamics if  $\Phi > 1$ .

### 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

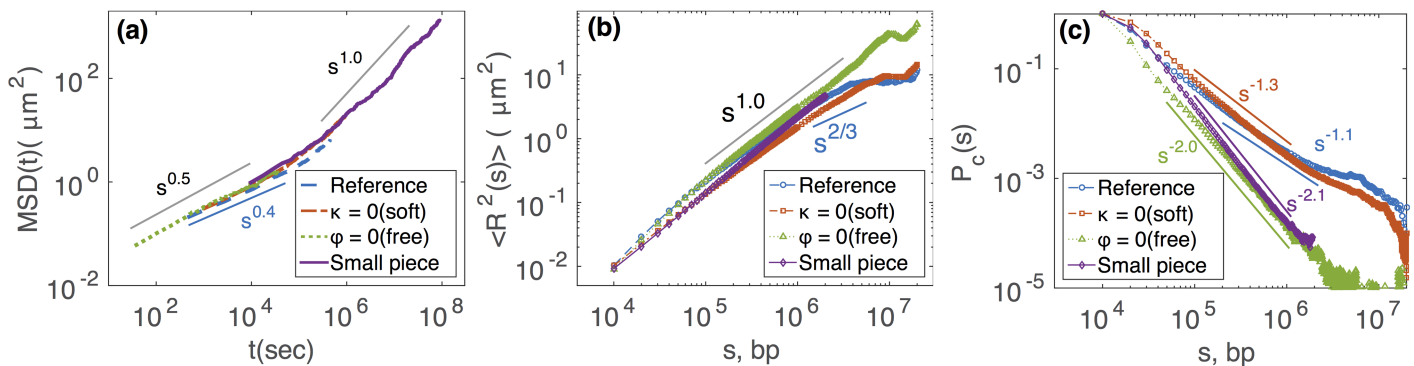


**Fig 3.** CPU time (given in hours) required to simulate 30 min of real dynamics for the drosophila case on a 3.20 GHz processor, (a) as a function of coarse graining where Kuhn sizes are fixed at three different values and, (b) as a function of Kuhn size at three different coarse grainings. The reference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

<https://doi.org/10.1371/journal.pcbi.1006159.g003>

coarse-graining? As we observed above, regarding numerical efficiency, one wants to go for the higher CG and higher  $\Phi$  ( $\lesssim 1$ ). These two values will determine the spatial and temporal resolutions of the model. Therefore, a natural choice would be to maximize CG and  $\Phi$  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  $\Phi$  so that we preserve the physical properties of the system by conserving the right entanglement regime. Now the question is what happens if one uses more naive coarse-graining strategies that do not necessarily preserve the topological regime. As explained at the beginning of section ‘Relation between Kuhn length  $l_k$  and lattice parameters’, 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 may modify the  $L/L_e$  ratio and therefore may 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 and 4c) (see also Fig.K bottom and Fig.L in S1 Text). In particular, considering an isolated chain ( $\Phi \ll 1$ ) dramatically



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

<https://doi.org/10.1371/journal.pcbi.1006159.g004>

modifies the behavior of  $P_c$  that scales within this approximation as  $\sim s^{-2}$ , characteristics of isolated self-avoiding walks [46]. This leads to an underestimation of the contact probability by orders of magnitude compared to the reference model.

In complement to coarse-graining strategies, still for 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 a small fragment of 2 Mbp, we found strong discrepancies. At small length scales,  $s < 100$  kbp, it follows the reference model, but at larger length scale  $100 \text{ kbp} < s < 2 \text{ Mbp}$  it deviates from reference model and behaves like an isolated Rouse polymer (see Fig 4 and Fig.M in S1 Text).

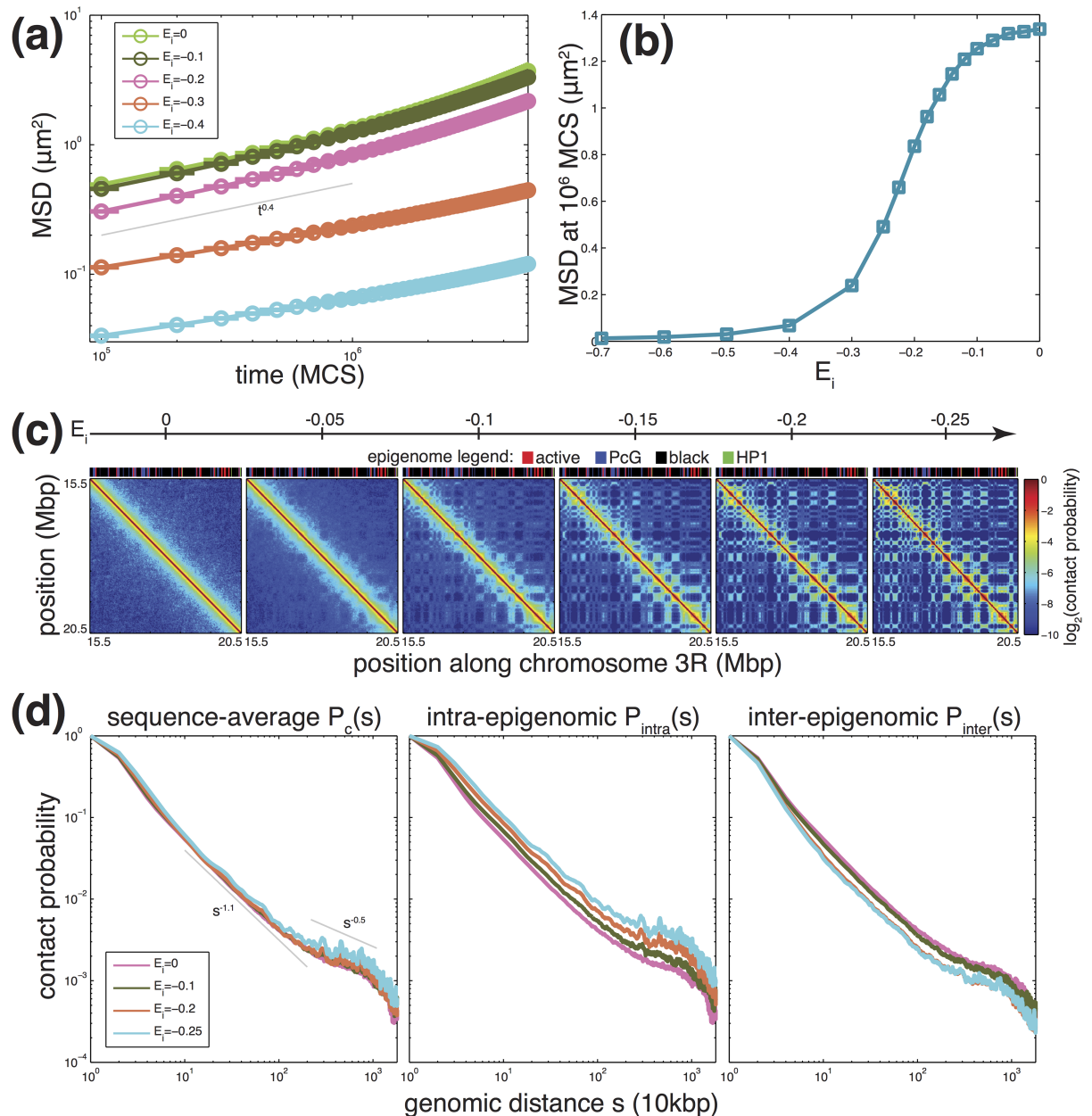
All this emphasizes the need to conserve properly the ratio  $L/L_e$  of the reference model if one aims to simulate the right polymeric behavior accurately. Modifying this ratio by decreasing  $L$  or by making approximations that affect  $L_e$  would possibly lead to simulate a system with different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

### 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, 61, 62]: 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 [63–65], 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 a 1D epigenomic domain. By varying the strength of these specific interactions, we showed that such model well accounts for the TAD formation and inter-TAD long-range contacts. Previously, we limited our analysis to short pieces of chromatin ( $\sim 1$  Mbp-long fragment) at equilibrium. In the previous section, 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  $\Phi = 0.97$  for  $L/L_e = 70$ ,  $\rho = 0.009 \text{ bp/nm}^3$ , see Table 1). For this region, we collect the epigenomic domains obtained by Filion et al [66] 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: constitutive 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  $E_i$  (in  $k_B T$  unit) if they are spatially in contact (ie nearest-neighbor on the lattice)(see Materials and methods). For simplicity, we assume that the strength of interaction is similar for every epigenomic state.

**Effect of varying the strength of specific interactions.** We first concentrate on the polymer dynamics by studying the average MSD along the simulations for various values of  $E_i$



**Fig 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  $k_B T$ -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 represent scaling laws.

<https://doi.org/10.1371/journal.pcbi.1006159.g005>

(Fig 5a). For all investigated interaction energies, the scaling properties of  $g_1$  are compatible with the diffusion of a crumpled polymer as seen in section ‘Generic behavior of chromosome’ with  $g_1 \sim t^{0.35-0.4}$ . As we increase the absolute value of 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  $E_i$  (Fig 5b), we observe a transition between a “fast” ( $E_i > -0.25$ ) and a “slow”

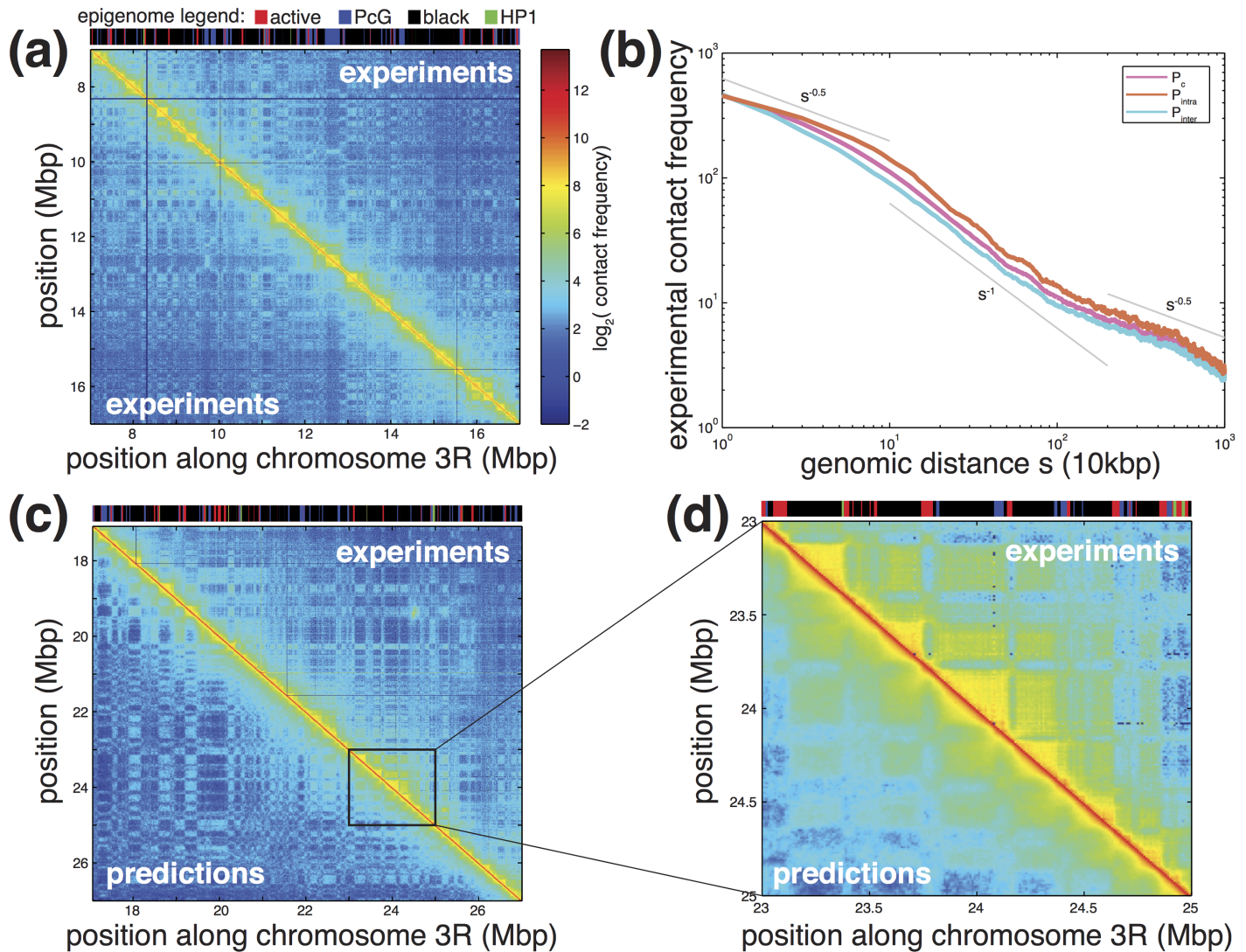
( $E_i < -0.25$ ) regime. This suggests a glass-like [67, 68] dynamic transition that occurs for strong specific interactions, reducing significantly the monomer mobility in the simulations.

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 ( $d_c = 163 \text{ nm}$ ), representing the average inside a population of unsynchronized cells with a typical cell cycle of 20 hrs [69]. 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 section ‘Generic behavior of chromosome’. As  $E_i$  is 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 [70]. 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 augment, leading to also an augmentation in  $P_c$ .

**Comparison with experimental data.** We next compare our results to Hi-C data obtained by Sexton et al for late drosophila embryos [61]. Experimental data exhibit the characteristic presence of TADs along the diagonal of the Hi-C map and of preferential long-range contacts between some TADs (Fig 6a and Fig.N in S1 Text). The sequence-average probability  $P_c(s)$  shows different regimes (Fig 6b): for  $s < 100 \text{ kbp}$ ,  $P_c(s) \sim s^{-0.5}$ , for  $100 \text{ kbp} < s < 1 \text{ Mbp}$ ,  $P_c(s) \sim s^{-1}$ , for  $1 \text{ Mbp} < s < 10 \text{ Mbp}$ ,  $P_c(s) \sim s^{-0.5}$ . 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 ( $s < 100 \text{ kbp}$ ) 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  $100 \text{ kbp} < s < 1 \text{ Mbp}$  and  $P_c(s) \sim s^{-0.5}$ , for  $1 \text{ Mbp} < s < 10 \text{ Mbp}$ . Among the different strength of specific interactions that we investigated,  $E_i = -0.1$  offers the best match with experimental data (Fig.O in S1 Text) 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 (Pearson correlation = 0.86) at the local—TAD—level but also at higher scales (Fig 6c and 6d and Fig.N in S1 Text), 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].

**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.

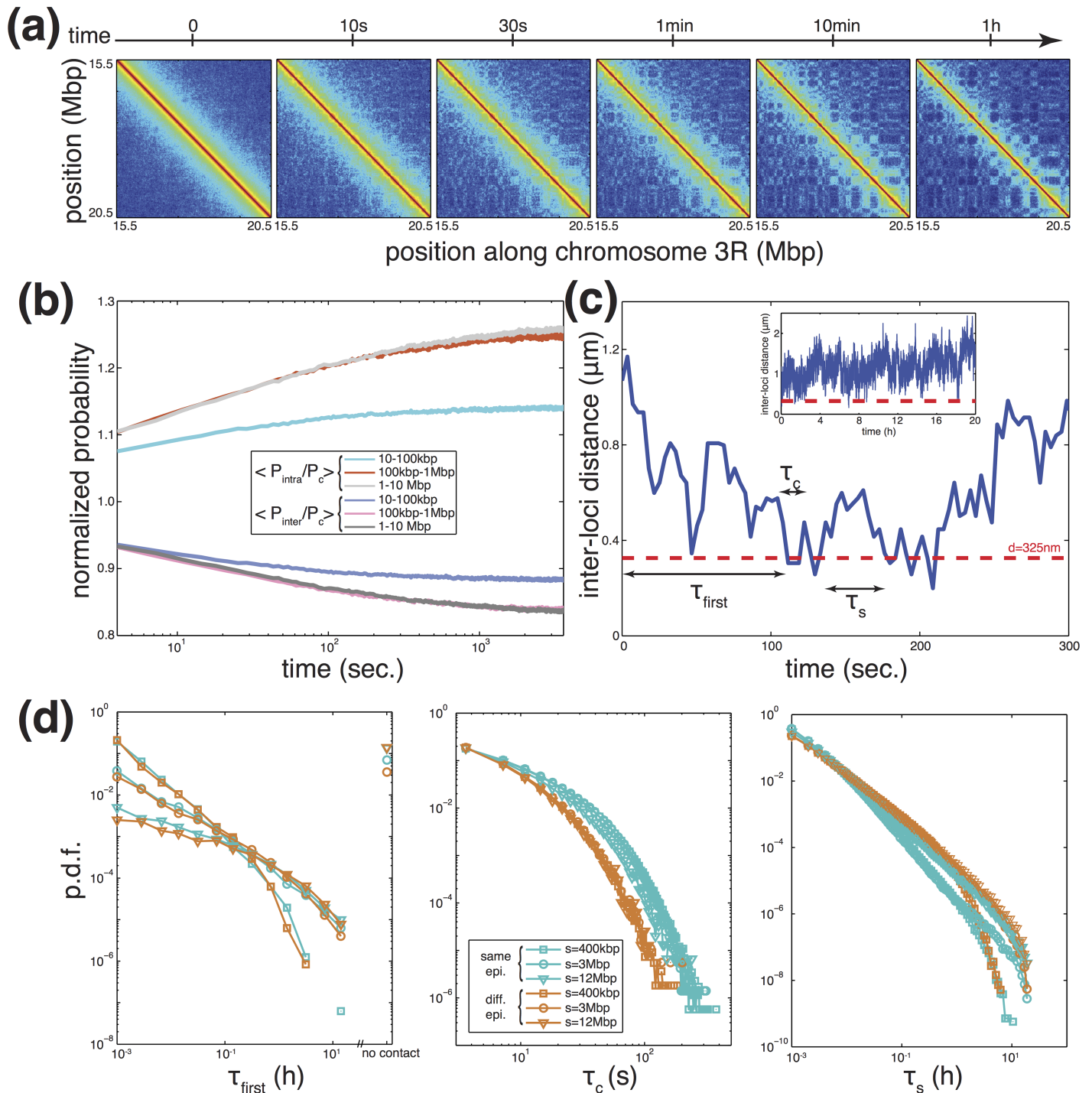


**Fig 6. Comparison between experimental and predicted data.** (a) Experimental Hi-C map for a 10 Mbp region of chromosome 3R. Corresponding epigenome is shown on top. (b) Average experimental contact frequency as a function of the genomic distance  $s$  between any pair of monomers (pink), between pairs of monomers having the same (orange) or different (cyan) epigenomic state. Grey lines represent 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.

<https://doi.org/10.1371/journal.pcbi.1006159.g006>

Time-evolution of the predicted Hi-C 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 reach steady-state very early in the cell cycle. For the 100 kbp–1 Mbp, convergence to 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  $\approx 3.5$  sec). Six pairs of loci having identical or different epigenomic states (Table C) and separated by different genomic lengths  $s$



**Fig 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  $\tau_{first}$  and several values for the contact time  $\tau_c$  and the search time  $\tau_s$ . (d) Probability distribution functions (p.d.f) of  $\tau_{first}$  (left),  $\tau_c$  (center) and  $\tau_s$  (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).

<https://doi.org/10.1371/journal.pcbi.1006159.g007>

( $s \sim 400$  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  $\tau_{first}$  defined as the first time after the mitotic exit when the pair becomes closer than  $d = 325$  nm; the contact time  $\tau_c$  defined as the time the pair stays in “contact” (ie closer than  $d$ ); and the search time  $\tau_s$  defined as the time interval between two “contacts”. For each pair, the probability distribution function of  $\tau_{first}$  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  $\tau_{first}$  are significantly observed. Interestingly, for  $s > 1$  Mbp, there 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  $\tau_c$  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  $\tau_s$  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  $\tau_s$ , 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  $\tau_s$ , memory is lost and the time interval between two contacts relies only on the genomic distance as for  $\tau_{first}$ .

## Discussion

In the first part of this article, we introduced a new coarse-graining strategy for long and dense self-avoiding walks that conserves the entanglement regime and the volumic density. Using kinetic Monte Carlo simulations on a lattice, we demonstrated that such strategy not only leads to the accurate description of 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 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 a big fraction of chromosome 3R of drosophila. It is the first study trying to quantitatively describe the behavior and time evolution of such large genomic regions (20 Mbp) during one cell cycle (20h of real time) with high precision (10 kbp resolution). Our heteropolymer model has one free 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  $\gamma$  depends on  $E_i$  and is sharply reduced above a given strength of interaction, characteristics of a glass-like transition [67, 68]. 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 [71]. 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, 41, 43, 58], 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 Hi-C 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 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 Hi-C or super-resolution experiments [72–77]. 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 the presence of cis-interacting mechanisms, like supercoiling [23, 78] or 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 Hi-C data made on synchronized mammalian cells showing that, in early G1, TADs are already observable in the data [76, 79]. Formation of long-range contacts require more time and eventually appear hours after the mitotic exit, also consistent with the time evolution of Hi-C data during the cell cycle in human [76, 79] and yeast [80]. 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  $\sim 10$  seconds, indicating a very dynamic situation, consistent with many experiments performed on living cells [81]. 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 [82]. 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 [83], 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 [84], 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 adequately 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  $\sim 120\text{h}$  of CPU time with our strategy at 10kbp resolution to simulate one cell cycle while it remains elusive 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, 85] 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 [86].

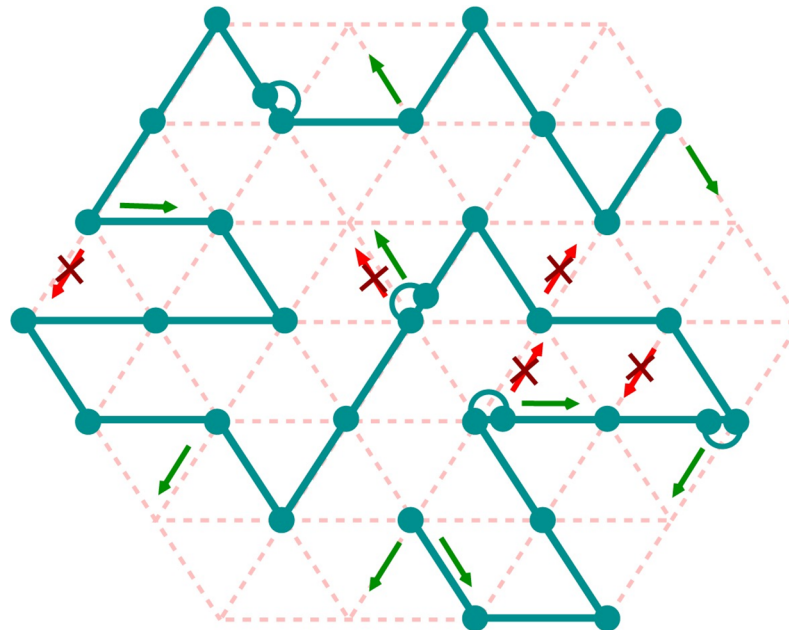
## Materials and methods

### Estimation of the chromatin volumic density

The chromatin volumic density  $\rho$  is defined as the ratio between the genome size  $G$  and the average volume  $V$  of the nucleus. For haploid yeast,  $G = 12.2\text{Mbp}$  and  $V \approx 2.6\mu\text{m}^3$  [69, 87], leading to  $\rho \approx 0.005 \text{ bp/nm}^3$ . For drosophila (diploid) late embryos, nuclei have a typical diameter of  $4\mu\text{m}$  [88] and contains about  $300\text{Mbp}$  of genomic DNA, thus  $\rho \approx 0.009 \text{ bp/nm}^3$ . In mammals, the size of the nucleus depends strongly on the cell type typically ranging from 5 to  $15\mu\text{m}$  in diameter [69]. For example, for a human nucleus ( $G \approx 6\text{Gbp}$ ) of diameter  $9\mu\text{m}$ ,  $\rho \approx 0.015 \text{ bp/nm}^3$ . For larger nuclei ( $12\mu\text{m}$  in diameter), as measured by Muller et al [89],  $\rho$  may be weaker ( $\approx 0.007 \text{ bp/nm}^3$ ).

### Simulation of structural and dynamical properties of the null model

The polymer is modeled as a semi-flexible self-avoiding walk, consisting of  $N$  beads, on a face centered cubic (fcc) lattice of size  $S \times S \times S$  (each unit cell contains 4 lattice sites) following the model developed by Hugouvieux et al [90] (Fig 8) (more details can be found in [32, 90, 91]). As in the elastic lattice model introduced recently by Schram and Barkema [92], we authorize at maximum two monomers to occupy the same lattice site if and only if they are consecutive along the chain [90, 92] (Fig 8). Otherwise, due to excluded volume, two non-consecutive monomers cannot be located at the same site. When two successive monomers occupy a lattice site, an extra bond length is accumulated in that node as a 'stored length' [92–94] (Fig. P(b) in S1 Text). The concept of stored length was first introduced by Rubinstein in his pioneering article on the implementation of repton model on a lattice [93]. Such double occupancy of consecutive monomers accounts for the effect of contour length fluctuations [94].



**Fig 8. Schematic of a lattice polymer configuration on a 2D projection of a 3D fcc lattice.** Solid line with beads represents the polymer chain, and the dotted lines represent the underlying lattice. Each lattice site is allowed to contain a maximum of two beads if and only 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.

<https://doi.org/10.1371/journal.pcbi.1006159.g008>

Rigidity is accounted using a standard formulation [95]:

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

where  $\kappa$  is the bending rigidity and is directly related to  $l_k/b$  (see below), and  $\theta_i$  is the angle between the bond vectors  $i$  and  $i + 1$ . Confinement and effect of other chains are approximated using periodic boundary conditions, the corresponding lattice volume fraction being  $\Phi = N/(4S^3)$ . Note that such periodic conditions do not confine the chain to the finite volume of the simulation box. Using correct unfolded coordinates, chains can extend over any large distances.

The dynamics of the chain follows a kinetic Monte-Carlo (KMC) scheme with simple local moves [90]: one Monte Carlo step (MCS) consists of  $N$  trial moves where a monomer is randomly chosen and randomly displaced to a nearest-neighbor site on the lattice (Fig 8). Trial moves are accepted according to a Metropolis criterion applied to  $\mathcal{H}$  and if the chain connectivity is maintained. Compared to standard Monte Carlo methods used to study systems at thermal equilibrium [96], KMC has the advantage to track the equilibrium or out-of-equilibrium dynamical evolution of a system. The transition rates in the KMC are assumed to be Poissonian which is likely to be an approximation of the exact dynamics at the relevant time scales. However, the connection between the real-time and KMC steps can be established precisely within the framework of Poissonian processes [97]. In our model, due to the approximated transition rates, accuracy is not guaranteed for time-scales below few MCS (temporal resolution) and length-scales below few  $b$  (spatial resolution).

This KMC scheme coupled to the notion of stored length (see above) allows 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 [92, 93]. Due to the simplicity and efficiency of such frameworks, they have been widely used in the literature to investigate various properties of polymeric systems [32, 90–92, 94, 98–100].

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 knot-free structures [16, 43, 101]. Therefore, we initiate our simulations by knot-free configurations generated using the ‘hedgehog’ algorithm [59, 102, 103]: 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.A(a) in S1 Text). We verified that, starting from other types of unknotted configurations such as Rosette (Fig.A(e) in S1 Text) and Cylindrical (Fig.A(f) in S1 Text), we recovered the same scaling laws for the null model (Fig.C in S1 Text).

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  $d_c$ . Note that all such properties are estimated using the ‘unfolded’ polymer conformations.

### 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  $\rho_{FS}$  leads to

$$\frac{L}{L_e} = \frac{Nb}{l_K} \left( \frac{\rho_{FS} b l_k^2}{cn} \right)^2 \tag{5}$$

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

$$\left( \frac{l_K}{b} \right)^3 = \left( \frac{n^3 c^2}{\rho_{FS}^2 b^6} \right) \left( \frac{b_0}{L_e^0} \right), \tag{7}$$

$\rho_{FS} = (N_0/V) (= \rho/v_0$  in the context of chromosome) with  $V$  the volume of the box. Conservation of the volume implies that

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

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

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

Practically, knowing  $\rho_{FS}$ ,  $b_0$  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 a relation between  $l_k/b$  (which is related to the bending energy of the model, see below) and the lattice volumic fraction  $\Phi$ . Eq 8 is used to compute the corresponding value for  $b$ .

### Relation between Kuhn size $N_\kappa$ and bending rigidity $\kappa$ 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 [104]

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

where

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

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

where  $\theta$  is the angle between two monomers and  $\kappa$  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  $N_\kappa$  to  $\kappa$ .

### Simulation of the block copolymer model for drosophila

In the block copolymer model, the energy of a given configuration is given by

$$H = \frac{\kappa}{2} \sum_{m=1}^{N-1} (1 - \cos \theta_m) + \sum_{l,m} U_{e(l),e(m)} \delta_{l,m}. \tag{13}$$

The first contribution accounts for the null model. The second contribution accounts for epigenomic-driven interactions with  $\delta_{l,m} = 1$  if monomers  $l$  and  $m$  occupy nearest-neighbor (NN) sites on the lattice ( $\delta_{l,m} = 0$  otherwise),  $e(l)$  the epigenomic state of monomer  $l$  and  $U_{e,e'}$  the strength of interaction between a pair of spatially neighbor beads of epigenomic states  $e$  and  $e'$ . For simplicity, we assume that interactions occur only between monomers of the same chromatin state ( $U_{e,e'} = 0$  if  $e \neq e'$ ) and that the strength of interaction (that we note  $E_i$ ) is the same whatever the chromatin state ( $U_{e,e} \equiv E_i$  for all  $e$ ). Dynamics of the chain follows the same KMC scheme as the null model using a Metropolis criterion applied to  $H$ . For various values of  $E_i$ , we simulate 400 trajectories during  $10^7$  MCS starting from a random unknotted “hedgehog” configuration (as in Fig.A(a,c) in S1 Text) at a standard *in vivo* bp-density ( $\rho = 0.009\text{bp/nm}^3$ ) (see the Movie for example). Note that such initial configurations might not reflect exactly the post-mitotic organization of fly chromosome and may impact the very large-scale—out-of-equilibrium—behavior predicted by the model.

## Supporting information

**S1 Text. A single pdf file containing 16 supporting figures and 3 supplementary tables.**  
(PDF)

**S1 Video. Time evolution of random unknotted ‘hedgehog’ configuration with specific interactions.**  
(AVI)

## Acknowledgments

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.

## Author Contributions

**Conceptualization:** Surya K. Ghosh, Daniel Jost.

**Data curation:** Surya K. Ghosh, Daniel Jost.

**Formal analysis:** Surya K. Ghosh, Daniel Jost.

**Funding acquisition:** Daniel Jost.

**Investigation:** Surya K. Ghosh, Daniel Jost.

**Methodology:** Surya K. Ghosh, Daniel Jost.

**Project administration:** Daniel Jost.

**Resources:** Surya K. Ghosh, Daniel Jost.

**Software:** Surya K. Ghosh, Daniel Jost.

**Supervision:** Daniel Jost.

**Validation:** Surya K. Ghosh, Daniel Jost.

**Visualization:** Surya K. Ghosh, Daniel Jost.

**Writing – original draft:** Surya K. Ghosh, Daniel Jost.

**Writing – review & editing:** Surya K. Ghosh, Daniel Jost.

## References

1. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015; 161(5):1012–1025. <https://doi.org/10.1016/j.cell.2015.04.004> PMID: 25959774
2. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*. 2016; 538:265–269. <https://doi.org/10.1038/nature19800> PMID: 27706140
3. Jost D, Vaillant C, Meister P. Coupling 1D modifications and 3D nuclear organization: data, models and function. *Curr Opin Cell Biol*. 2017; 44:20–27. <https://doi.org/10.1016/j.ceb.2016.12.001> PMID: 28040646
4. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*. 2001; 2(4):292–301. <https://doi.org/10.1038/35066075> PMID: 11283701
5. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009; 326(5950):289–293. <https://doi.org/10.1126/science.1181369> PMID: 19815776

6. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014; 159(7):1665–80. <https://doi.org/10.1016/j.cell.2014.11.021> PMID: 25497547
7. Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, et al. Comparative analysis of metazoan chromatin organization. *Nature*. 2014; 512(7515):449–52. <https://doi.org/10.1038/nature13415> PMID: 25164756
8. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012; 485(7398):376–380. <https://doi.org/10.1038/nature11082> PMID: 22495300
9. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012; 485(7398):381–385. <https://doi.org/10.1038/nature11049> PMID: 22495304
10. Zhan Y, Mariani L, Barozzi I, Schulz ED, Blüthgen N, Stadler M, 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 Research*. 2017; 27:479–490. <https://doi.org/10.1101/gr.212803.116> PMID: 28057745
11. Ulianov SV, Khrameeva EE, Gavrillov AA, Flyamer IM, Kos P, Mikhaleva EA, et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res*. 2016; 26(1):70–84. <https://doi.org/10.1101/gr.196006.115> PMID: 26518482
12. Zhu Y, Chen Z, Zhang K, Wang M, Medovoy D, Whitaker JW, et al. Constructing 3D interaction maps from 1D epigenomes. *Nat Commun*. 2016; 7:10812. <https://doi.org/10.1038/ncomms10812> PMID: 26960733
13. Mourad R, Cuvier O. Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation. *PLoS Computational Biology*. 2016; 12:e1004908. <https://doi.org/10.1371/journal.pcbi.1004908> PMID: 27203237
14. Haddad N, Vaillant C, Jost D. IC-Finder: inferring robustly the hierarchical organization of chromatin folding. *Nucleic Acids Res*. 2017; 45:e81. <https://doi.org/10.1093/nar/gkx036> PMID: 28130423
15. Sexton T, Cavalli G. The role of chromosome domains in shaping the functional genome. *Cell*. 2015; 160(6):1049–59. <https://doi.org/10.1016/j.cell.2015.02.040> PMID: 25768903
16. Rosa A, Everaers R. Structure and Dynamics of Interphase Chromosomes. *PLOS Computational Biology*. 2008; 4(8):1–10. <https://doi.org/10.1371/journal.pcbi.1000153>
17. Hyeon C, Thirumalai D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nature Communications*. 2011; 2(487). <https://doi.org/10.1038/ncomms1481> PMID: 21952221
18. Barbieri M, Chotalia M, Fraser J, Lavitas LM, Dostie J, Pombo A, et al. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A*. 2012; 109(40):16173–16178. <https://doi.org/10.1073/pnas.1204799109> PMID: 22988072
19. Sharma R, Jost D, Kind J, Gómez-Saldivar G, van Steensel B, Askjaer P, et al. Differential spatial and structural organization of the X chromosome underlies dosage compensation in *C. elegans*. *Genes Dev*. 2014; 28(23):2591–2596. <https://doi.org/10.1101/gad.248864.114> PMID: 25452271
20. Jost D, Carrivain P, Cavalli G, Vaillant C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*. 2014; 42(15):9553. <https://doi.org/10.1093/nar/gku698> PMID: 25092923
21. Tark-Dame M, Jerabek H, Manders EMM, Heermann DW, 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*. 2014; 10(10):e1003877. <https://doi.org/10.1371/journal.pcbi.1003877> PMID: 25299688
22. Ganai N, Sengupta S, Menon GI. Chromosome positioning from activity-based segregation. *Nucleic Acids Res*. 2014; 42(7):4145–4159. <https://doi.org/10.1093/nar/gkt1417> PMID: 24459132
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*. 2014; 42(5):2848–2855. <https://doi.org/10.1093/nar/gkt1353> PMID: 24366878
24. Giorgetti L, Galupa R, Nora EP, Pilot T, Lam F, Dekker J, et al. Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription. *Cell*. 2014; 157:950–963. <https://doi.org/10.1016/j.cell.2014.03.025> PMID: 24813616
25. Jerabek H, Heermann DW. How chromatin looping and nuclear envelope attachment affect genome organization in eukaryotic cell nuclei. *Int Rev Cell Mol Biol*. 2014; 307:351–381. <https://doi.org/10.1016/B978-0-12-800046-5.00010-2> PMID: 24380599
26. Sanborn AL, Rao SSP, Huang SC, Durand NC, Huntley MH, Jewett AI, 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. 2015; 112(47):E6456–E6465. <https://doi.org/10.1073/pnas.1518552112> PMID: 26499245
27. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*. 2016; 15(9):2038–2049. <https://doi.org/10.1016/j.celrep.2016.04.085> PMID: 27210764
  28. Goloborodko A, Imakaev MV, Marko JF, Mirny L. Compaction and segregation of sister chromatids via active loop extrusion. *Elife*. 2016; 5. <https://doi.org/10.7554/eLife.14864> PMID: 27192037
  29. Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*. 2016; 529(7586):418–422. <https://doi.org/10.1038/nature16496> PMID: 26760202
  30. Pierro MD, Zhang B, Aiden EL, Wolynes PG, Onuchic JN. Transferable model for chromosome architecture. *Proc Natl Acad Sci U S A*. 2016; 113:12168–12173. <https://doi.org/10.1073/pnas.1613607113> PMID: 27688758
  31. Wani AH, Boettiger AN, Schorderet P, Ergun A, Munger C, Sadreyev RI, et al. Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat Commun*. 2016; 7:10291. <https://doi.org/10.1038/ncomms10291> PMID: 26759081
  32. Olarte-Plata JD, Haddad N, Vaillant C, Jost D. The folding landscape of the epigenome. *Physical Biology*. 2016; 13(2):026001. <https://doi.org/10.1088/1478-3975/13/2/026001> PMID: 27042992
  33. Chiariello AM, Annunziatella C, Bianco S, Esposito A, Nicodemi M. Polymer physics of chromosome large-scale 3D organisation. *Sci Rep*. 2016; 6:29775. <https://doi.org/10.1038/srep29775> PMID: 27405443
  34. Brackley CA, Johnson J, Kelly S, Cook PR, Marenduzzo D. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res*. 2016; 44(8):3503–3512. <https://doi.org/10.1093/nar/gkw135> PMID: 27060145
  35. Brackley CA, Brown JM, Waithe D, Babbs C, Davies J, Hughes JR, et al. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biology*. 2016; 17:59. <https://doi.org/10.1186/s13059-016-0909-0> PMID: 27036497
  36. Tiana G, Amitai A, Pollex T, Piolot T, Holcman D, Heard E, et al. Structural Fluctuations of the Chromatin Fiber within Topologically Associating Domains. *Biophys J*. 2016; 110(6):1234–1245. <https://doi.org/10.1016/j.bpj.2016.02.003> PMID: 27028634
  37. Florescu AM, Therizols P, Rosa A. Large Scale Chromosome Folding Is Stable against Local Changes in Chromatin Structure. *Plos Comput Biol*. 2016; 12(6):e1004987. <https://doi.org/10.1371/journal.pcbi.1004987> PMID: 27295501
  38. Iyer BVS, Arya G. Lattice animal model of chromosome organization. *Phys Rev E*. 2012; 86:011911. <https://doi.org/10.1103/PhysRevE.86.011911>
  39. Meluzzi D, Arya G. Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Research*. 2013; 41(1):63–75. <https://doi.org/10.1093/nar/gks1029> PMID: 23143266
  40. Hyeon C, Thirumalai D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nature Communications*. 2011; 2:487. <https://doi.org/10.1038/ncomms1481> PMID: 21952221
  41. Halverson JD, Smrek J, Kremer K, Grosberg AY. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Rep Prog Phys*. 2014; 77(2):022601. <https://doi.org/10.1088/0034-4885/77/2/022601> PMID: 24472896
  42. Cremer T, Markaki Y, Hübner B, Zunhammer A, Strickfaden H, Beichmanis S, et al. Chromosome Territory Organization within the Nucleus. *Reviews in Cell Biology and Molecular Medicine*. 2012;.
  43. Mirny LA. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res*. 2011; 19(1):37–51. <https://doi.org/10.1007/s10577-010-9177-0> PMID: 21274616
  44. Putz M, Kremer K, Grest GS. What is the Entanglement Length in a Polymer Melt? *Europhys Lett*. 2000; 49:735. <https://doi.org/10.1209/epl/i2000-00212-8>
  45. Uchida N, Grest GS, Everaers R. Viscoelasticity and primitive path analysis of entangled polymer liquids: From F-actin to polyethylene. *The Journal of Chemical Physics*. 2008; 128(4):044902. <https://doi.org/10.1063/1.2825597> PMID: 18247995
  46. De Gennes PG. Scaling concepts in polymer physics. Cornell University Press. 1980; 32(5):290–290.
  47. Hajjoul H, Mathon J, Ranchon H, Goiffon I, Mozziconacci J, Albert B, et al. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Research*. 2013; 23(11):1829–1838. <https://doi.org/10.1101/gr.157008.113> PMID: 24077391
  48. Munkel C, Langowski J. Chromosome structure predicted by a polymer model. *Phys Rev E*. 1998; 57:5888–5896. <https://doi.org/10.1103/PhysRevE.57.5888>



49. Cheutin T, Cavalli G. Progressive polycomb assembly on H3K27me3 compartments generates polycomb bodies with developmentally regulated motion. *Plos Genet.* 2012; 8(1):e1002465. <https://doi.org/10.1371/journal.pgen.1002465> PMID: 22275876
50. Doi M, Edwards SF. *The Theory of Polymer Dynamics.* Clarendon Press. 1988; 32(5):290–290.
51. Tamm MV, Nazarov LI, Gavrilov AA, Chertovich AV. Anomalous Diffusion in Fractal Globules. *Phys Rev Lett.* 2015; 114:178102. <https://doi.org/10.1103/PhysRevLett.114.178102> PMID: 25978267
52. Farge E, Maggs AC. Dynamic scattering from semiflexible polymers. *Macromolecules.* 1993; 26(19):5041–5044. <https://doi.org/10.1021/ma00071a009>
53. Bystricky K, Heun P, Gehlen L, Langowski J, Gasser SM. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc Natl Acad Sci USA.* 2004; 101:16495–16500. <https://doi.org/10.1073/pnas.0402766101> PMID: 15545610
54. Kimura H, Shimooka Y, Nishikawa Ji, Miura O, Sugiyama S, Yamada S, et al. The genome folding mechanism in yeast. *J Biochem.* 2013; 154(2):137–147. <https://doi.org/10.1093/jb/mvt033> PMID: 23620598
55. Duan Z, Andronescu M, Schutz K, Mclwain S, Kim YJ, Lee C, et al. A three-dimensional model of the yeast genome. *Nature.* 2010; 465:363–367. <https://doi.org/10.1038/nature08973> PMID: 20436457
56. Grosberg A, Rabin Y, Havlin S, Neer A. Crumpled globule model of the three-dimensional structure of DNA. *Europhys Lett.* 1993; 23:373–378. <https://doi.org/10.1209/0295-5075/23/5/012>
57. Rosa A, Becker NB, Everaers R. Looping probabilities in model interphase chromosomes. *Biophys J.* 2010; 98:2410–2419. <https://doi.org/10.1016/j.bpj.2010.01.054> PMID: 20513384
58. Rosa A, Everaers R. Ring polymers in the melt state: the physics of crumpling. *Phys Rev Lett.* 2014; 112:118302. <https://doi.org/10.1103/PhysRevLett.112.118302> PMID: 24702424
59. Imakaev MV, Tchourine KM, Nechaev SK, Mirny LA. Effects of topological constraints on globular polymers. *Soft Matter.* 2015; 11:665–671. <https://doi.org/10.1039/c4sm02099e> PMID: 25472862
60. Lucas JS, Zhang Y, Dudko OK, Murre C. 3D Trajectories Adopted by Coding and Regulatory DNA Elements: First-Passage Times for Genomic Interactions. *Cell.* 2014; 158:339–352. <https://doi.org/10.1016/j.cell.2014.05.036> PMID: 24998931
61. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell.* 2012; 148(3):458–472. <https://doi.org/10.1016/j.cell.2012.01.010> PMID: 22265598
62. Haddad N, Jost D, Vaillant C. Perspectives: using polymer modeling to understand the formation and function of nuclear compartments. *Chromosome Research.* 2017; 25(1):35–50. <https://doi.org/10.1007/s10577-016-9548-2> PMID: 28091870
63. Canzio D, Liao M, Naber N, Pate E, Larson A, Wu S, et al. A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature.* 2013; 496(7445):377–381. <https://doi.org/10.1038/nature12032> PMID: 23485968
64. Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, et al. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell.* 2013; 26(6):565–577. <https://doi.org/10.1016/j.devcel.2013.08.016> PMID: 24091011
65. Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, et al. Liquid droplet formation by HP1 suggests a role for phase separation in heterochromatin. *Nature.* 2017;. <https://doi.org/10.1038/nature22822>
66. Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell.* 2010; 143(2):212–224. <https://doi.org/10.1016/j.cell.2010.09.009> PMID: 20888037
67. Donth EJ. *The Glass Transition: Relaxation Dynamics in Liquids and Disordered Materials.* Springer-Verlag; 2001.
68. Mezard M, Parisi G, Virasoro M. *Spin glass theory and beyond.* World Scientific, Singapore; 1987.
69. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 2010; 38(Database issue):D750–3. <https://doi.org/10.1093/nar/gkp889> PMID: 19854939
70. Miles IS, Rostami S, editors. *Multicomponent Polymer Systems.* Longman Scientific and Technica, Singapore; 1992.
71. Gürsoy G, Xu Y, Kenter AL, Liang J. Spatial confinement is a major determinant of the folding landscape of human chromosomes. *Nucleic Acids Research.* 2014; 42(13):8223–8230. <https://doi.org/10.1093/nar/gku462> PMID: 24990374

72. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 2013; 502(7469):59–64. <https://doi.org/10.1038/nature12593> PMID: 24067610
73. Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Ting Wu C, et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. 2016; 353:598–602. <https://doi.org/10.1126/science.aaf8084> PMID: 27445307
74. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*. 2017; 544:59–64. <https://doi.org/10.1038/nature21429> PMID: 28289288
75. Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. 2017; 544:110–114. <https://doi.org/10.1038/nature21711> PMID: 28355183
76. Nagano T, Lubling Y, Varnai C, Dudley C, Leung W, Baran Y, et al. Cell cycle dynamics of chromosomal organisation at single-cell resolution. *Nature*. 2017; 547:61–67. <https://doi.org/10.1038/nature23001> PMID: 28682332
77. Cattoni DI, Cardozo-Gizzi AM, Georgieva M, Stefano MD, Valeri A, Chamousset D, et al. Single-cell absolute contact probability detection reveals that chromosomes are organized by modulated stochasticity. *bioRxiv*. 2017;.
78. Racko D, Benedetti F, Dorier J, Stasiak A. Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. *Nucleic Acids Research*. 2017; p. gxk1123.
79. Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, et al. Organization of the Mitotic Chromosome. *Science*. 2013; 342(6161):948–953. <https://doi.org/10.1126/science.1236083> PMID: 24200812
80. Lazar-Stefanita L, Scolari V, Mercy G, Thierry A, Muller H, Mozziconacci J, et al. Choreography of budding yeast chromosomes during the cell cycle. *bioRxiv*. 2017;.
81. Bystricky K. Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. *FEBS Letters*. 2015; 589:3014–3022. <https://doi.org/10.1016/j.febslet.2015.07.012> PMID: 26188544
82. Polovnikov KE, Gherardi M, Cosentino-Lagomarsino M, Tamm MV. Folding and cytoplasm viscoelasticity contribute jointly to chromosome dynamics. *arXiv*. 2017; p. arXiv:1703.10841.
83. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*. 2014; 32:347–355. <https://doi.org/10.1038/nbt.2842> PMID: 24584096
84. Chen H, Fujioka M, Gregor T. Direct visualization of transcriptional activation by physical enhancer-promoter proximity. *bioRxiv*. 2017;.
85. Mattout A, Cabianca DS, Gasser SM. Chromatin states and nuclear organization in development—a view from the nuclear lamina. *Genome Biology*. 2015; 16:174. <https://doi.org/10.1186/s13059-015-0747-5> PMID: 26303512
86. Chandra T, Ewels PA, Schoenfelder S, Furlan-Magaril M, Wingett SW, Kirschner K, et al. Global Reorganization of the Nuclear Landscape in Senescent Cells. *Cell Rep*. 2015; 10:471–483. <https://doi.org/10.1016/j.celrep.2014.12.055> PMID: 25640177
87. Jorgensen P, Edgington NP, Schneider BL, Rupeš I, Tyers M, Fitcher B. The Size of the Nucleus Increases as Yeast Cells Grow. *Molecular Biology of the Cell*. 2007; 18:3523–3532. <https://doi.org/10.1091/mbc.E06-10-0973> PMID: 17596521
88. Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, et al. Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell*. 2011; 144(2):214–226. <https://doi.org/10.1016/j.cell.2010.12.026> PMID: 21241892
89. Muller I, Boyle S, Singer RH, Bickmore WA, Chubb JR. Stable Morphology, but Dynamic Internal Reorganisation, of Interphase Human Chromosomes in Living Cells. *PLoS One*. 2010; 5:e11560. <https://doi.org/10.1371/journal.pone.0011560> PMID: 20644634
90. Hugouvieux V, Axelos MAV, Kolb M. Amphiphilic Multiblock Copolymers: From Intramolecular Pearl Necklace to Layered Structures. *Macromolecules*. 2009; 42:392–400. <https://doi.org/10.1021/ma801337a>
91. Szabo Q, Jost D, Chang JM, Cattoni DI, Papadopoulos GL, Bonev B, et al. TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Science Advances*. 2018; 4(2). <https://doi.org/10.1126/sciadv.aar8082> PMID: 29503869
92. Schram RD, Barkema GT. Simulation of ring polymer melts with GPU acceleration. *Journal of Computational Physics*. 2018; 363:128–139. <https://doi.org/10.1016/j.jcp.2018.02.027>

93. Rubinstein M. Discretized model of entangled-polymer dynamics. *Phys Rev Lett.* 1987; 59:1946–1949. <https://doi.org/10.1103/PhysRevLett.59.1946> PMID: 10035375
94. Newman MEJ, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. *Phys Rev E.* 2001; 64:026118. <https://doi.org/10.1103/PhysRevE.64.026118>
95. Auhl R, Everaers R, Grest GS, Kremer K, Plimpton SJ. Equilibration of long chain polymer melts in computer simulations. *The Journal of Chemical Physics.* 2003; 119(24):12718–12728. <https://doi.org/10.1063/1.1628670>
96. Binder K, Heermann D W. Monte Carlo simulation in statistical physics: an introduction. Springer. 2002;.
97. Fichtthorn KA, Weinberg WH. Theoretical foundations of dynamical Monte Carlo simulations. *The Journal of Chemical Physics.* 1991; 95(2):1090–1096. <https://doi.org/10.1063/1.461138>
98. van Heukelum A, Beljaars HRW. Electrophoresis simulated with the cage model for reptation. *The Journal of Chemical Physics.* 2000; 113(9):3909–3915. <https://doi.org/10.1063/1.1287784>
99. Wolterink JK, Barkema GT. Polymer diffusion in a lattice polymer model with an intrinsic reptation mechanism. *Molecular Physics.* 2005; 103(21-23):3083–3089. <https://doi.org/10.1080/00268970500208682>
100. Jost D, Vaillant C. Epigenomics in 3D: importance of long-range spreading and specific interactions in epigenomic maintenance. *Nucleic Acids Research.* 2018; 46:2252–2264. <https://doi.org/10.1093/nar/gky009> PMID: 29365171
101. Zhang B, Wolynes PG. Topology, structures, and energy landscapes of human chromosomes. *Proc Natl Acad Sci U S A.* 2015; 112:6062–6067. <https://doi.org/10.1073/pnas.1506257112> PMID: 25918364
102. Earnshaw WC, Laemmli UK. Architecture of metaphase chromosomes and chromosome scaffolds. *The Journal of Cell Biology.* 1983; 96(1):84–93. <https://doi.org/10.1083/jcb.96.1.84> PMID: 6826654
103. Klenin KV, Vologodskii AV, Anshelevich VV, Dykhne AM, Frank-Kamenetskii MD. Effect of Excluded Volume on Topological Properties of Circular DNA. *Journal of Biomolecular Structure and Dynamics.* 1988; 5(6):1173–1185. <https://doi.org/10.1080/07391102.1988.10506462> PMID: 3271506
104. Doi M, Edwards SF. *The Theory of Polymer Dynamics.* Oxford University Press, USA; 1986.