Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains

Daniel Jost[1], Pascal Carrivain[2], Giacomo Cavalli[2],* and Cédric Vaillant[1],*

[1]Laboratoire de Physique, Ecole Normale Supérieure de Lyon, CNRS UMR 5672, Lyon 69007, France and [2]Institute of Human Genetics, CNRS UPR 1142, Montpellier 34000, France

ABSTRACT

Genomes of eukaryotes are partitioned into domains of functionally distinct chromatin states. These domains are stably inherited across many cell generations and can be remodeled in response to developmental and external cues, hence contributing to the robustness and plasticity of expression patterns and cell phenotypes. Remarkably, recent studies indicate that these 1D epigenomic domains tend to fold into 3D topologically associated domains forming specialized nuclear chromatin compartments. However, the general mechanisms behind such compartmentalization including the contribution of epigenetic regulation remain unclear. Here, we address the question of the coupling between chromatin folding and epigenome. Using polymer physics, we analyze the properties of a block copolymer model that accounts for local epigenomic information. Considering copolymers build from the epigenomic landscape of *Drosophila,* we observe a very good agreement with the folding patterns observed in chromosome conformation capture experiments. Moreover, this model provides a physical basis for the existence of multistability in epigenome folding at sub-chromosomal scale. We show how experiments are fully consistent with multistable conformations where topologically associated domains of the same epigenomic state interact dynamically with each other. Our approach provides a general framework to improve our understanding of chromatin folding during cell cycle and differentiation and its relation to epigenetics.

INTRODUCTION

Gene expression is regulated by many sets of proteins that associate with the genome in a cell-type and condition-specific manner at specific regulatory elements including proximal promoters, enhancers and repressors. The packaging of eukaryotic DNA into chromatin contributes to this regulation via the modulation of the accessibility and specificity of regulators to their nuclei sites. Locally, the chromatin state is characterized by various features like the nucleosome positioning, the covalent modifications of DNA and histones tails and the insertion of histone variants. This pattern of chromatin states along the genome, the so-called 'epigenome', is itself regulated by the combined action of different specialized chromatin regulators like chromatin remodelers, modifying enzymes and histone chaperones.

The general picture that emerges from the genome-wide high-resolution profiling of structural and functional chromatin marks obtained in various organisms and cell types (1–4), is that eukaryotic genomes are linearly organized into distinct epigenomic domains. These domains extend over few kilobases up to few megabases, are characterized by a specific type of chromatin and are isolated from their neighborhood by boundary elements such as insulators. Euchromatin, less condensed, early replicating and containing most active genes, is generally distinguished from heterochromatin, typically highly condensed, late replicating and inhibitory to transcriptional machinery. In many higher eukaryotes, from plants to mammals, statistical analyses of hundreds of chromatin marks have identified only a small number of main chromatin types (1,3,5,6), typically four or five, covering the well-known constitutive HP1-like heterochromatin or the facultative (developmentally regulated) Polycomb-like heterochromatin but also a less-characterized ultra-repressive heterochromatin enriched in genes that are expressed in very few tissues, the so-called void or black chromatin (1,7).

Interestingly, within epigenomic domains, regulatory sequences such as enhancers may be located far from the target genes and multiple elements that are distributed over large regions may collaborate or compete for the regulation of individual genes or gene clusters. This implies the existence of long-range mechanisms where regulatory elements could act over large genomic distances up to hundreds of kilobases or more. A possible mechanism regulating such long-range effects is the linear spreading of a

*To whom correspondence should be addressed. Tel: +33 4 72 72 86 34; Fax: +33 4 72 72 89 50; Email: cedric.vaillant@ens-lyon.fr
Correspondence may also be addressed to Giacomo Cavalli. Tel: +33 4 34 35 99 70; Fax: +33 4 34 35 99 01; Email: giacomo.cavalli@igh.cnrs.fr

regulatory signal (e.g. repressive chromatin state) from nucleation sites (e.g. silencers) to target-sites (e.g. promoters). Another non-exclusive mechanism calls into play the polymeric nature of chromatin that may induce spatial colocalization of regulatory sequences with their target. Recently, chromosome conformation capture (3C)-based studies have indeed shown that regulatory elements can act over large genomic distances by chromatin looping (8,9) forming active or repressive higher-order chromatin structure at particular developmentally regulated genes. These pairwise 3D interactions are mediated by DNA binding proteins such as insulators or cohesin and mediator (10) that would cluster in space and bridge distant regulatory sites. At a genomic scale, the contact maps of *Drosophila* (11,12), mouse (13) and human (13,14) chromosomes have further revealed a remarkable 3D compartmentalization where epigenomic domains fold into independent 'spatial domains', the so-called topologically associated domains (TADs), characterized by (i) high intra-domain contact frequencies; (ii) 3D insulation between adjacent domains; (iii) and in many cases, significant contacts between distal domains of the same chromatin type (Figures 3A and 4A). This compartmentalization is consistent with the nuclear structure, as revealed by imaging techniques such as electron microscopy and immuno-FISH (15–17), that clearly shows a phase separation between euchromatin versus heterochromatin and to some extent between the different heterochromatin types (18).

Altogether, these experimental data reveal a clear link between the epigenome, the 3D chromosome architecture and the gene transcription pattern. However, it is still unclear what are the mechanisms behind 1D epigenome assembly and its 3D folding in TADs, and how these 1D and 3D organizations precisely contribute to gene regulation. Here, we address the question of the 3D folding of the epigenome using physical modeling. In particular, we ask if it is possible to interpret the observed correlations between the epigenome and the 3D chromatin organization in TADs by means of polymer physics arguments.

A generic and minimal model to investigate the large-scale 3D organization and dynamics of chromosomes is to consider chromatin as a semi-flexible self-avoiding homopolymer (19). The conformation of such polymer is controlled by the interplay between thermal motion, steric repulsion and effective monomer–monomer interactions that may account for steric confinement, attractive or repulsive interactions between monomers or interactions mediated by the solvent. For example, in a 'good solvent' condition, when steric repulsion dominates, the chain adopts a swollen coil state (20). In a 'poor solvent' condition, when effective attraction prevails, the polymer chain becomes compressed and equilibrates into a collapsed globular phase characterized by a high monomer density and a contact frequency that is almost constant at large separation distance. Between these two conditions, at the transition point (the so-called θ-temperature) where repulsion and attraction counterbalance, the chain adopts a Gaussian coil state. Such homopolymeric model with appropriate geometric constraints provide a fairly good description of large-scale conformational properties from yeast to human and in particular a good fit of different scaling behaviors like the dependence of the contact frequency or of the mean distance on the linear genomic distance. In yeast, static and dynamic studies suggest a brush-like equilibrated organization (21,22) of chromosomes with scaling properties compatible with a weakly collapsed state; in fly and human, where chromosomes are much longer, large-scale organization is characterized by a dense state that has been primarily associated to a non-equilibrium fractal globule composed of spatially segregated long-lived domains (23,24) but that can also be mapped to a semi-dilute solution of non-concatenated rings at equilibrium (25,26) or to a linear homopolymer with dynamic attractive self-interactions (27,28).

However, most of these models consider a homopolymer that cannot obviously account for the regional variability of the chromatin organization and in particular for the spatial compartmentalization of the epigenome. One has to introduce genomic and/or epigenomic specificities in the folding model (27,29–32). Here, we posit that chromatin folding is driven by effective epigenomic-dependent interactions between chromatin loci. To test this hypothesis, we propose a generic and new theoretical approach by treating chromatin as a block copolymer, where each block corresponds to an epigenomic domain and where each monomer interacts preferentially with other monomers of the same chromatin type. This is largely motivated by the observations of self-interactions between chromatin types (11) and is also supported by increasing evidence that some architectural proteins might promote physical bridging (33,34). We adapt a Gaussian self-consistent approach (35,36) to derive equilibrium phase diagram and contact maps for any block copolymers, as a function of two control parameters, the specific and non-specific monomer–monomer interaction. When considering block copolymers build from the epigenomic landscape of *Drosophila* (1), we show that this simple physical model accounts very well for the folding patterns in TADs observed in Hi-C experiments (11). As a main and very original outcome, this model provides a physical basis for the existence of multistability in chromosome organization. We show indeed how some experimental patterns are fully consistent with multistable conformations where TADs of the same epigenomic state interact transiently or long-lastly with each other.

MATERIALS AND METHODS

Block copolymer model

Chromatin is modeled as a self-avoiding bead-spring polymer containing N monomers, each monomer representing 10 kb of DNA. A given conformation of the chain is characterized by a Hamiltonian H. The Hamiltonian is made of two contributions $H = H_{chain} + H_{inter}$ where H_{chain} describes the self-avoiding Gaussian chain and H_{inter} accounts for attractive short-range interactions between monomers. Connectivity between successive beads is modeled by a spring and the self-avoidance by a repulsive hard-code potential. The self-avoiding Gaussian chain is then described by

$$H_{chain} = \frac{3k_B T}{2l^2} \sum_n (X_n - X_{n-1})^2 + \sum_{n<m} U_{hc}(r_{nm}) \quad (1)$$

with X_n the position of monomer n, l the bond length of the pure Gaussian chain, r_{nm} the distance between monomers

n and m, and $U_{\text{hc}}(r_{nm})$ a truncated Lennard–Jones like potential (see Supplementary Notes). Contact interactions between monomers are modeled by a Gaussian-like potential:

$$H_{\text{inter}} = \sum_{n<m} E_{nm} \exp\left[-\frac{r_{nm}^2}{2r_0^2}\right] \quad (2)$$

with E_{nm} the strength of interaction between monomers n and m, r_0 defining the length-range of the interaction. We note that the effects described in the article are robust over different parameter values, Hamiltonian forms and size of the investigated region (see Supplementary Figures S1 and S2).

Gaussian self-consistent approximation

To simplify, the stochastic dynamics of the chain is modeled by a set of coupled Langevin equations (37):

$$\xi \frac{d\boldsymbol{X}_n}{dt} = -\frac{\partial H}{\partial \boldsymbol{X}_n} + \boldsymbol{\eta}_n(t), \quad n = 1, \dots, N \quad (3)$$

with ξ the friction coefficient and $\boldsymbol{\eta}_n$ delta-correlated white noises that accounts for stochastic fluctuations of the system. The probability distribution function of $Y = \{\boldsymbol{X}_n\}$ verifies therefore the Fokker–Planck equation (38):

$$\frac{\partial P}{\partial t} = \frac{1}{\xi} \sum_n \left[\frac{\partial}{\partial \boldsymbol{X}_n} \left(P \frac{\partial H}{\partial \boldsymbol{X}_n} \right) + k_{\text{B}} T \frac{\partial^2 P}{\partial \boldsymbol{X}_n^2} \right] \quad (4)$$

At each time point, we approximate P by a multivariate Gaussian distribution $P(Y, t) \approx (1/Z)\exp[-Y^+ C(t)^{-1} Y/2]$ with $C(t) = \{\langle \boldsymbol{X}_n \cdot \boldsymbol{X}_m \rangle /3\}$ the covariance matrix of Y at time t. To find an evolution equation for C, we adapt the approach developed by Ramalho *et al.* for biochemical reaction networks (36) to polymer dynamics: starting from the Gaussian distribution at time t, we can estimate with Equation (1) the evolved distribution at $t + \delta t$, then we look for the Gaussian distribution that better describes this evolved distribution in terms of relative entropy, Kullback–Leibler divergence or Gibbs free energy. This leads to (see Supplementary Notes)

$$\xi \frac{dC}{dt} = \langle J \rangle C + C \langle J^+ \rangle + N_{\text{r}} \quad (5)$$

with N_{r} the covariance matrix of the random processes $\{\boldsymbol{\eta}_n\}$ and $\langle J \rangle = -\int dY P(Y, t) (\partial^2 H)/(\partial \boldsymbol{X}_n \partial \boldsymbol{X}_m)$, the average value (over the current Gaussian distribution) of the Hessian matrix of H. Since, by definition, the mean squared distance between monomer m and n is given by $D_{mn} = \langle (\boldsymbol{X}_m - \boldsymbol{X}_n)^2 \rangle /3 = C_{mm} + C_{nn} - 2C_{mn}$ and $\langle J \rangle$ is a function of $\{D_{mn}\}$ (see Supplementary Notes), Equation (5) leads directly to the self-consistent equation:

$$\xi \frac{dD_{mn}}{dt} = 4k_{\text{B}} T$$
$$- \sum_k (\langle J_{mk} \rangle - \langle J_{nk} \rangle)(D_{mk} - D_{nk}) \quad (6)$$

where $k_{\text{B}} T$ is the typical amplitude of stochastic fluctuations. Note that Timoshenko *et al.* have derived a similar
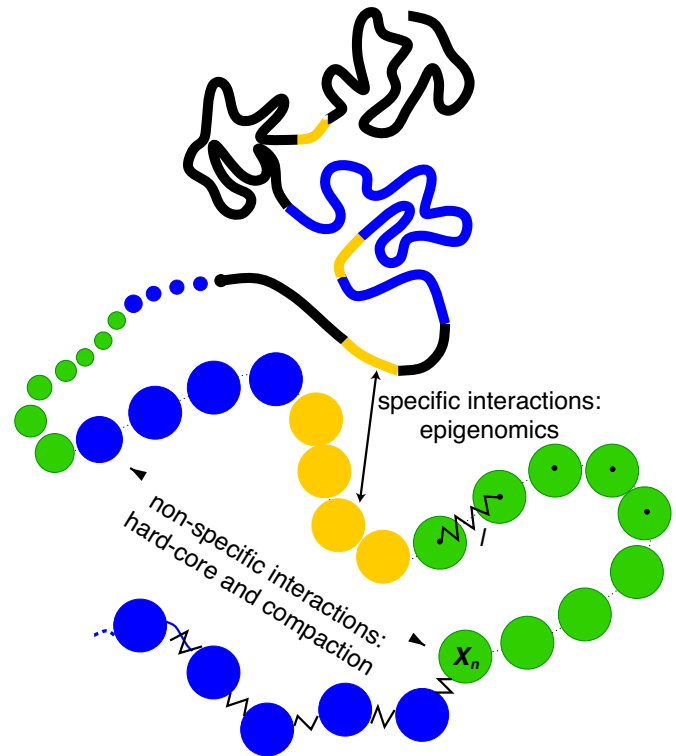
Figure 1. Block copolymer model: the chromatin is modeled as a self-avoiding bead-spring chain where each monomer represents a portion of DNA (10 kb) and is characterized by its epigenetic state: yellow (active), green (HP1-like heterochromatin), blue (Polycomb-like heterochromatin), black (repressive chromatin) (1). The model integrates non-specific and specific short-range interactions to account respectively for the effective compaction of the chain and for epigenomically related affinities between monomers.

equation for copolymers using Gibbs–Bogoliubov inequality (35).

Equation (6) is solved by numerical integration: starting from different initial conditions, we use a fifth order adaptive Runge–Kutta algorithm (39) until a stationary regime is achieved.

Numerical simulations

Full numerical simulations of the block copolymer model were performed using a home-made molecular dynamics program. Integration of the trajectories was implemented using a standard velocity-Verlet algorithm and thermalization of the system was carried out by an Andersen thermostat (40) (see Supplementary Notes). Starting from a random configuration, we first let the system reach equilibrium before taking measurements on the system. We note that the results on the dynamics of the multistability region are robust over different stochastic collision frequencies of the thermostat (see Supplementary Figure S3).

RESULTS

Modeling chromatin as a block copolymer

The 1D sequence of epigenomic domains along the polymer chromatin allows the mapping between the system of

interest and a block copolymer. Block copolymers are heteropolymers composed by successive blocks of different monomers. They have been intensively studied in the fields of polymer physics and chemistry and exhibit many remarkable properties, like their ability to form structured spatial conformations very similar to chromatin organization in the nucleus. In this paper, we propose a minimal extension of the homopolymeric model by considering chromatin as a block copolymer with interaction terms that would depend on the local epigenomic state. This introduces immiscibility between the monomers of different types that is well known to induce phase separation (20).

We model the chromatin fiber as an interacting self-avoiding bead-and-spring chain. Each monomer represents a portion of DNA (10 kb) and is characterized by an epigenetic state (see Figure 1). In addition to standard excluded volume interactions, we consider attractive short-range interactions E_{mn} between monomers. In the following, we assume only two types of interactions: (i) non-specific interactions U_{ns} between every pairs of monomers that effectively account for compaction effect due to confinement into the nucleus; and (ii) specific attractive interactions U_s between monomers having the same epigenetic state (Figure 1). Hence $E_{mn} = U_{ns} + \delta_{mn} U_s$ with $\delta_{mn} = 1$ (resp. 0) if monomers m and n have (resp. have not) the same epigenetic state.

This specificity is motivated by many experimental evidence suggesting effective interactions between loci of identical chromatin state. Indeed, it has been shown that Polycomb group (34,41,42) or HP1 (33) proteins may create physical bridges between distant heterochromatin regions. Furthermore, mutualization of transcription machinery resources or DNA looping mediated by architectural proteins like cohesin and mediator may also lead to effective attractions between active loci (9,10,12,43,44). And, black chromatin is often associated with lamins (1) suggesting effective interactions mediated by the nuclear lamina. For simplicity, we assume that all the specific interactions have the same strengths (i.e. for example two Polycomb-like monomers interact with the same intensities than two HP1-like monomers).

While numerical simulations of such stochastic system could be time demanding, we develop a Gaussian self-consistent approximation that allows a fast scanning of the parameter space and that enables the efficient computation of expectation values for the HiC-map of the copolymer. This type of approach has already proven to be useful and accurate when studying copolymers (35,45). At each time point, the probability distribution function of the chain conformation is approximated by a multivariate Gaussian distribution whose parameters are determined using the maximum entropy principle (36) (see 'Material and Methods' section for details). This approximation leads to a closed set of ordinary differential equations that describe the dynamics of the mean squared distances between monomers (Equation (6)). We solve this set of non-linear equations in the steady-state limit by numerical integration. From the computed squared distances D_{mn}, we estimate the probability of contact P_{mn} between two monomers using the Gaussian approximation $P_{mn} \approx A D_{mn}^{-3/2}$ (with A a constant numerical factor). It has to be noted that the self-

consistent approach can only give an approximate solution of chromatin folding due to the rough treatment of excluded volume. To verify that our conclusions are not artifacts of the approximation, we also perform for some parameter sets full numerical simulations of the copolymer using molecular dynamics (see 'Material and Methods' section). Moreover, we make the assumptions that the polymer chains are equilibrated. Detailed simulations of confined self-avoiding homopolymers (24) have suggested that strong topological constraints may slow down the chain dynamics such that equilibration of very long confined chains can be extremely long (24), well above the cell cycle length. However, recent theoretical studies (26) have estimated that below few Mb, the typical length scale investigated in this work, topological confinement is negligible and polymer chains may safely be considered as equilibrated.

Heterogeneous chromatin exhibits a complex phase diagram with multistability

To illustrate the generic effects predicted by the model, we consider, as a toy example, a chain of 120 beads with an alternation of active (A) and black (B) epigenetic domains of the same size $((A_{10}B_{10})_6)$. Figure 2 shows the richness and complexity of the observed behaviors as we vary the strengths of compaction (via U_{ns}) and of specificity (via U_s) even for such simple epigenetic sequence. The phase diagram is made of four different regions. For weak compaction and specificity, the system is in a coil phase with extended chain conformations. As we increase the compaction at weak specificity, the chain undergoes a θ-collapse transition to a globular compact phase. For strong compaction and specificity, we observe checkerboard-like contact map characteristics of microphase separated (MPS) conformations where all monomers of the same epigenetic state are densely packed into distinct 3D domains. This phase is closed to the intermingled phase observed by Jerabek and Heermann when they introduce a sinusoidal binding affinity in their dynamic loop model (30). Between the coil and MPS phases, lies a region of multistability where Equation (6) has multiple fixed points depending on the initial conditions. Theses solutions are metastable intermediate configurations between coil and MPS. For example, heat-map d1 in Figure 2 represents pearl-necklace conformations where epigenetic domains have internally collapsed but remain isolated from each other, forming topologically associated domains. The size and location of the multistability region depend on the properties of the copolymer like the number of blocks, the number of block types, or the linear organization of the blocks along the polymer. However, we qualitatively observe that the size of the region grows with the complexity of the epigenomic sequence (Supplementary Figure S4). The enlargement of the area goes often with an increase of the number of metastable states (35).

To go further in the characterization of the multistability, we perform full numerical simulations for sets of parameters inside the region of interest of the phase diagram. We observe that in most conformations the epigenomic domains are internally collapsed (Figure 2B, left). The dynamics of the chain is then composed by stochastic jumps between several families of metastable states (Figure 2B, right). Each
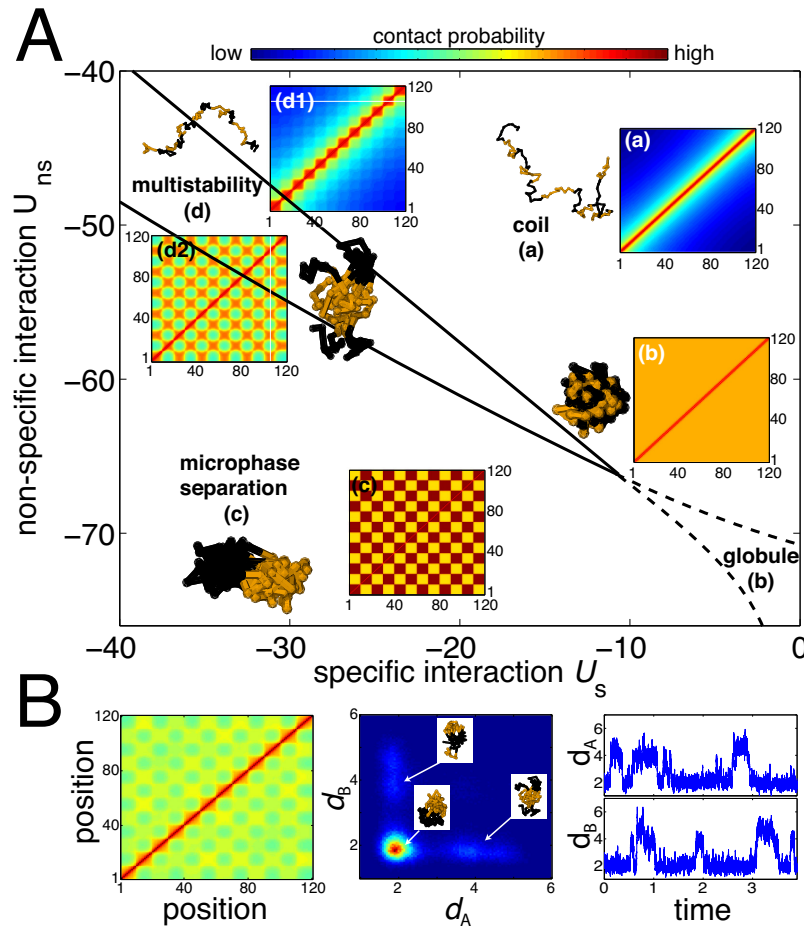
Figure 2. (**A**) Phase diagram of the copolymer $(A_{10}B_{10})_6$ as a function of the strength of specific and non-specific interactions (in k_BT unit). Insets represent typical heat maps of the probability of contacts between two monomers (in log-unit) for the different phases: coil (a), globule (b), multiphase separation (c) and multistability (d). Snapshots result from full numerical simulations of the system. (**B**) Contact map (left), joint-probability distribution function for the root mean squared distance (r.m.s.d.) d_A between A-monomers and the r.m.s.d. d_B between B-monomers (center), and typical time-evolution of d_A and d_B along one simulated trajectory (right), obtained from full numerical simulations for a parameter set inside the multistability region (d_2 in (A)). Time is given in arbitrary simulation time-unit (see Supplementary Notes).

family being characterized by transient contacts between two or more epigenomic domains of the same type that temporarily merge together (Figure 2B, center). Association and dissociation dynamics between domains depend on the position inside the multistability region with faster dynamics close to the boundary with the coil phase (Supplementary Figure S3).

Drosophila **Hi-C maps are compatible with multistability**

In this section, we apply the block copolymer model to different chromatin regions of *Drosophila melanogaster*. These regions have been chosen to be representative of the generic features of experimental Hi-C contact maps and of *in vivo* chromatin organization. Recently, Filion *et al.* have characterized the epigenome of embryonic *Drosophila* cell line Kc167 via statistical analysis of Dam-ID profiles of several chromatin-associated proteins (1). They found five typical epigenomic states: two euchromatic states that differ at the level of gene functions, and three heterochromatic states: the constitutive heterochromatin enriched in HP-1 proteins,

the facultative Polycomb heterochromatin enriched in genes implicated in differentiation and development, as well as the so-called black chromatin, the prevalent type of repressive chromatin. In the following, we will use the epigenomic data from Filion *et al.*, where for simplicity we merged the two active states into one single epigenomic type, as the primary sequence of copolymers and vary the non-specific and specific interaction strengths to draw phase diagram as in the toy example.

Microphase separation in black chromatin. We start with a chromatin region (located between 23.05 and 24.36 Mb of chromosome 3R) whose epigenetic state is composed of large black domains separated by short active domains (Figure 3A). Its experimental Hi-C map (Figure 3A) shows the internal folding of black and active domains with almost uniform intra-domain contact probabilities. Inter-domain contacts are numerous between black domains. We also observe long-range contacts between active domains. For example, the active region A_1 around 23.77 Mb is able to loop out the compact globule formed by the neighbor black do-
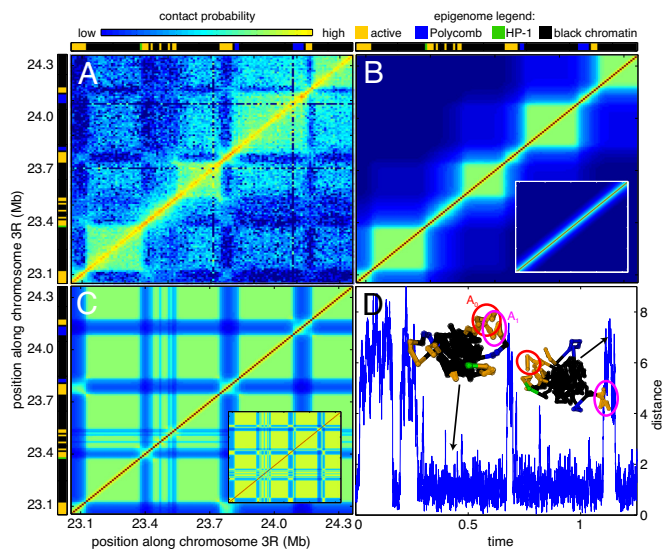
Figure 3. (A) Experimental Hi-C contact map for the chromatin region located between 23.05 and 24.36 Mb of chromosome 3R (from (11)). Epigenetic domains (from (1)) are given at the top and at the left borders of the figure: active (orange), Polycomb (blue), HP-1 (green) and black chromatin. (B and C) Examples of predicted contact maps inside the multistability region ($U_{ns} = -25\,k_BT$, $U_s = -63\,k_BT$) starting from a coil (B) or a MPS (C) configuration (see insets). (D) Time-evolution of the distance between the centers of masses of the active domains A_0 and A_1 along one simulated trajectory. Insets represent typical conformations of the chain.
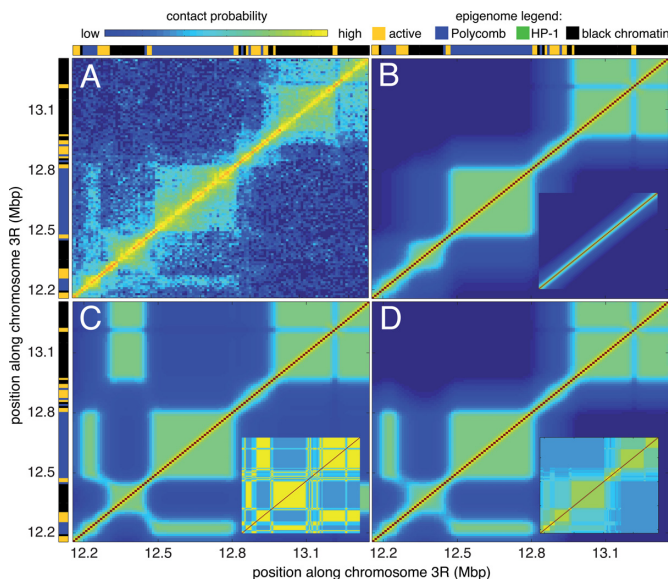
Figure 4. (A) Experimental Hi-C contact map for the chromatin region located between 12.16 and 13.36 Mb of chromosome 3R (from (11)). Epigenetic domains (from (1)) are given at the top and at the left borders of the figure: active (orange), Polycomb (blue), HP-1 (green) and black chromatin. (B, C and D) Examples of predicted contact maps inside the multistability region ($U_{ns} = -40\,k_BT$, $U_s = -44\,k_BT$) starting from a coil (B), a MPS (C) or a experimental-like (D) configuration (see insets).

mains to contact another active domain A_0 (around 23.10 Mb) located 600 kb apart. In the context of a copolymer, these experimental observations are consistent with MPS-like conformations. Using the self-consistent approxima-

tion, we study the phase diagram of the region (modeled by a chain of 131 beads with 10 kb per bead) (Supplementary Figure S4). Many stationary solutions exist, that are consistent with the experimental Hi-C map. They are all located within the multistability region (examples are given in Figure 3B and C). In particular, we were able to reproduce the pattern of inter-black domains interactions and some long-range contacts between active domains (Figure 3C). By scanning a small part of the multistability region using full simulations, we find sets of parameter that catch qualitatively the average polymeric behavior of the chain as well as the specificity of the experimental data (Supplementary Figure S5). Of note, this suggests that black chromatin forms a compact metastable globule that transiently dissociates, and that small active domains are expelled at the periphery of the globule (insets Figure 3D). This localization allows dynamic interactions between the active regions (Figure 3D).

Long-range contacts between Polycomb domains. As another example, we choose the chromatin region located between 12.16 and 13.36 Mb of chromosome 3R. The epigenetic state is composed of a succession of black and Polycomb (blue) domains separated by short active regions (Figure 4A). The Hi-C map (Figure 4A) is made of internally folded domains corresponding to the epigenetic domains. Of particular interest are the long-range contacts observed between the Polycomb domains centered around 12.23 Mb (P_1) and 12.65 Mb (P_2), and the looping of the small active region (around 13.22 Mb) out the dense globule of black chromatin (B_0) where it is embedded in. To investigate if the copolymer model is able to describe the experimental observations, we generate the phase diagram of this region (modeled by a chain of 120 beads) (Supplementary Figure S4). We observe the same four regions with a larger multistability area, consistent with the higher complexity of the local epigenome. In this region, depending on the initial conditions, many fixed points may be found for a fixed set of parameters, characteristics of frustrated phases that become dominant in copolymers with random sequences of monomers (35). The lack of information on chromatin organization in fly during mitosis does not allow to well defined the 'true' initial conditions. Like the previous black chromatin region, we observe that experiments are consistent with the multistability region. However, starting from coil, globular or MPS conformations, we were not able to reproduce the full characteristics of the experimental data (Supplementary Figure S6). While the internal folding of epigenetic domains and the looping out of the small active region are well described, we fail to mimic the inter-domain contacts, notably the simultaneous presence of contacts between P_1 and P_2 and absence of contacts between B_0 and the black domain (B_1) centered around 12.39 Mb. However, if the initial condition mimics the experimental Hi-C maps (see Supplementary Notes), within a significant portion of the multistability region, the system converges to a metastable state very close to the observed data (Figure 4D). This underlines the importance of initial conditions in the epigenome folding and suggests possible memory effects with the maintenance of long-range contacts already present in the mitotic chromatin organization. An alternative possibility to the mitotic

reminiscence of long-range interactions is the existence of heterogeneities in the strengths of specific interactions. Indeed, assuming that Polycomb–Polycomb interactions are stronger than intra-black–chromatin interactions also allows to recapitulate the experimental data (Supplementary Figure S7).

DISCUSSION

A simple theoretical framework for epigenome folding

In this article, we introduce a block copolymer model to investigate the folding properties of chromosomes as a function of the underlying epigenome. This model considers a single input, namely the experimentally derived epigenome that defines the primary sequence of the copolymer, and only two control parameters, the non-specific and the specific monomer–monomer interactions. The former accounts for a global compaction level, the latter for the effective attraction between monomers of same chromatin type. We used an efficient computational approach that allows us to derive expected contact maps and explore phase diagrams over a broad range of parameters. Remarkably, we show that such a minimal model can account for the main generic properties observed experimentally. Our approach provides a simple, tractable and attractive theoretical framework for interpreting the organization of the epigenome at sub-chromosomal scale (from 10 kb to few Mb) and in particular its spatial compartmentalization in TADs. Here, folding of the genome is assumed to be mainly driven by effective attractive interactions between chromatin elements of same epigenomic types. Self-association promotes internal folding and leads to spatial segregation and insulation of adjacent epigenomic domains without the need to introduce any bridging or anchoring activities at the TAD boundaries. Increasing specific attraction essentially leads to further compaction of the TADs and global confinement promotes cross-talk between TADs of the same epigenomic state. Compartmentalization might be a way of coordinating and reinforcing the functional output of genomic regions by colocalization and mutualization of the same specific regulators.

Chromatin organization is multistable and dynamic

One of the main outcomes of the copolymer model is the existence of a multistability region inside the phase diagram where epigenomic domains fold into topologically associated domains that interact transiently with each other. The dynamics of these interactions depend on the strength of the specific interactions, but also on the sizes of the epigenomic domains. For example, small domains, like most of the epigenomically active domains, would exhibit fast and dynamic interactions, while bigger domains may form long-lived metastable interactions. Comparison with experimental Hi-C maps of *Drosophila* suggests that biological situations are consistent with this multistability. This implies that, *in vivo*, chromatin organization is being dynamically and stochastically remodeled while conserving local key features like the TADs. This prediction of the model is in perfect agreement with recent single-cell Hi-C experiments in mouse showing the conservation of TADs be-
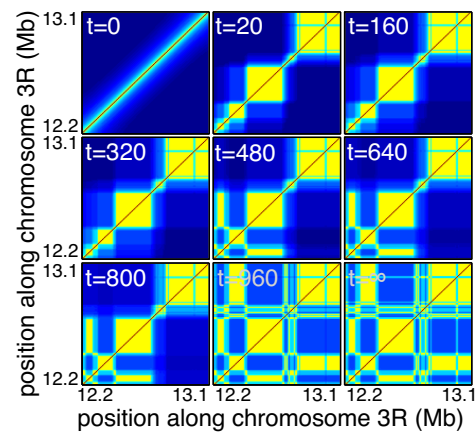
Figure 5. Evolution of the contact map for the chromatin region located between 12.16 and 13.36 Mb as a function of time *t* (in arbitrary simulation time-unit), starting from a coil-like conformation and ending at steady-state in a MPS-like conformation. Legend color as in Figure 4.

tween cells and the high variability of inter-domain contacts (14). Stochasticity in long-range inter-TAD interactions may represent an important source of intrinsic noise that may play an important role on the co-regulation of distant genes (46). However, during development or differentiation, since noisy systems closed to criticality are very sensitive to external stimuli (47,48), this variability might be useful to respond efficiently to developmental cues that drive the colocalization of distant loci (49).

Recent experiments performed on senescent cells have shown the nuclear rearrangement of heterochromatic marks into non-overlapping micro-domains (18). Within our formalism, this suggests that chromatin organization may relax to microphase separation configuration in non-dividing cell. Figure 5 shows the dynamic evolution of a contact map predicted by the copolymer model starting from a coil state and ending in a MPS steady-state. Interestingly, we observe the very fast formation of TADs, followed by a long period of slow compaction where long-range interactions are gradually incorporated, until the copolymer experiences a very fast transition to MPS. This intermediate slowing-down is a signature of the glassy-like dynamics of copolymers when crossing the multistability/frustrated region (35). These predictions are also consistent with recent Hi-C experiments on synchronized HeLa cells (50) showing that the formation of TADs is already achieved in early G1 starting from a mitotic conformation where the organization in TADs is apparently lost, and that the Hi-C map remains fairly unchanged throughout the cell cycle except during mitosis. This suggests that in normal dividing cells, chromatin organization converges quickly to multistable conformations and does not have the time to relax to a MPS-like state due to the periodic reinitialization of the chromatin organization at mitosis.

Toward inference and prediction

In this paper, we aimed at exploring the generic folding properties of the chromatin fiber. Therefore, for simplicity, we limit our approach to the simplest version of a block

copolymer that we can build from the compartmentalization of the epigenome. The model assumes that all specific interactions have the same strength whatever the chromatin type and whatever the location along the genome. However, such two-parameter model needs to be refined in order to gain in predictability. In particular, the current model considers each domain as a homopolymer with uniform monomer–monomer interactions. This can only provide a coarse-grained understanding of epigenome folding and cannot account for variation in contact frequency and in particular for preferential pairwise (long-range) contact between discrete genomic loci. Recent studies (9,12) have proposed that these site- and lineage-specific contacts mediated by architectural proteins (insulators, cohesin and mediators) might indeed play a key role in the folding of chromosomes at the sub-Mb scale. Along the same line, anchoring at the membrane of particular sequence or/and epigenomic domains (via their association with lamina or nuclear pores) has been shown to be crucial for organizing chromatin inside the nucleus (30,51,52). In addition to the global, non-site-specific, interactions investigated in this study, focal large-scale looping and anchoring might indeed contribute to spatial compartmentalization of domains (11,12,53). Therefore, further improvements of the model will require to augment the number of parameters by allowing for variability of interaction at the monomer scale, and to infer specific interaction strengths that predict at best the observed contact maps.

The copolymer framework associated with the self-consistent Gaussian approximation may represent an efficient formalism to extract from the available experimental data the effective genomic and epigenomic interactions between chromatin loci (54). As a promising outcome of such inference process, would be a powerful tool to predict the chromatin organization in various conditions, allowing investigating *in silico* changes in TAD formations and long-range contacts when altering the epigenome. In particular, during development, cell differentiation proceeds by global and concomitant rearrangements of epigenomic profile, chromatin organization and transcriptional activity (49,55–57). Hence, our model may provide a very interesting framework for understanding how epigenome regulation (resp. deregulation) during development (resp. disease) could lead to cell phenotypic variations via large-scale chromatin reorganization.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

We thank Ralf Everaers, Angelo Rosa and Tom Sexton for fruitful discussions. We also acknowledge the Pôle Scientifique de Modélisation Numérique for computing resources.

FUNDING

REFERENCES

1. Filion,G.J., van Bemmel,J.G., Braunschweig,U., Talhout,W., Kind,J., Ward,L.D., Brugman,W., de Castro,I.J., Kerkhove,R.M., Bussemake,H.J. *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in drosophila cells. *Cell*, **143**, 212–224.
2. Ernst,J., Kheradpour,P., Mikkelsen,T.S., Shoresh,N., Ward,L.D., Epstein,C.B., Zhang,X., Wang,L., Issner,R., Coyne,M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.
3. Roudier,F., Ahmed,I., Bérard,C., Sarazin,A., Mary-Huard,T., Cortijo,S., Bouyer,D., Caillieux,E., Duvernois-Berther,E., Al-Shikhley,L. *et al.* (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J.*, **30**, 1928–1938.
4. Gerstein,M.B., Lu,Z.J., van Nostrand,E.L., Cheng,C., Arshinoff,B.I., Liu,T., Yip,K.Y., Robilotto,R., Rechtsteiner,A., Ikegami,K. *et al.* (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science*, **330**, 1775–1787.
5. Kharchenko,P.V., Alekseyenko,A.A., Schwartz,Y.B., Minoda,A., Riddle,N.C., Ernst,J., Sabo,P.J., Larschan,E., Gorchakov,A.A.,, Gu,T. *et al.* (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature*, **471**, 480–485.
6. Dekker,J. (2007) GC- and AT-rich chromatin domains differ in conformation and histone modification status and are differentially modulated by rpd3p. *Genome Biol.*, **8**, R116.
7. van Steensel,B. (2011) Chromatin: constructing the big picture. *EMBO J.*, **30**, 1885–1895.
8. Ong,C.T. and Corces,V.G. (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.*, **12**, 283–293.
9. Phillips-Cremins,J.E., Sauria,M.E.G., Sanyal,A., Gerasimova,T.I., Lajoie,B.R., Bell,J.S.K., Ong,C.T., Hookway,T.A., Guo,C., Sun,Y. *et al.* (2013) Architectural protein subclasses shape 3d organization of genomes during lineage commitment. *Cell*, **153**, 1281–1295.
10. Sofueva,S., Yaffe,E., Chan,W.C., Georgopoulou,D., Rudan,M.V., Mira-Bontenbal,H., Pollard,S.M., Schroth,G.P., Tanay,A. and Hadjur,S. (2013) Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.*, **32**, 3119–3129.
11. Sexton,T., Yaffe,E., Kenigsberg,E., Bantignies,F., Leblanc,B., Hoichman,M., Parrinello,H., Tanay,A. and Cavalli,G. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, **148**, 458–472.
12. Hou,C., Li,L., Qin,Z.S. and Corces,V.G. (2012) Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell*, **48**, 471–484.
13. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
14. Nagano,T., Lubling,Y., Stevens,T.J., Schoenfelder,S., Yaffe,E., Dean,W., Laue,E.D., Tanay,A. and Fraser,P. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, **502**, 59–64.
15. Ahmed,K., Dehghani,H., Rugg-Gunn,P., Fussner,E., Rossant,J. and Bazett-Jones,D.P. (2010) Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS One* **5**, e10531.
16. Rapkin,L.M., Anchel,D.R.P., Li,R. and Bazett-Jones,D.P. (2012) A view of the chromatin landscape. *Micron* **43**, 150–158.
17. Solovei,I., Wang,A.S., Thanisch,K., Schmidt,C.S., Krebs,S., Zwerger,M., Cohen,T.V., Devys,D., Foisner,R., Peichl,L. *et al.* (2013) Lbr and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152**, 584–598.
18. Chandra,T., Kirschner,K., Thuret,J-Y, Pope,B.D., Ryba,T., Newman,S., Ahmed,K., Samarajiwa,S.A., Salama,R., Carroll,T. *et al.* (2012) Independence of repressive histone marks and chromatin

compaction during senescent heterochromatic layer formation. *Mol. Cell* **47**, 203–214.

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

20. Grosberg,A.Y. and Khokhlov,A.R. (1994) *Statistical Physics of Macromolecules*. AIP Press, Woodbury, NY.

21. Wong,H., Marie-Nelly,H., Herbert,S., Carrivain,P., Blanc,H., Koszul,R., Fabre,E. and Zimmer,C. (2012) A predictive computational model of the dynamic 3d interphase yeast nucleus. *Curr. Biol.* **22**, 1881–1890.

22. Hajjoul,H., Mathon,J., Ranchon,H., Goiffon,I., Mozziconacci,J., Albert,B., Carrivain,P., Victor,J.M., Gadal,O., Bystricky,K. *et al.* (2013) High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res.* **23**, 1829–1838.

23. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293.

24. Rosa,A. and Everaers,R. (2008) Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* **4**, e1000153.

25. Halverson,J.D., Lee,W.B., Grest,G.S., Grosberg,A.Y. and Kremer,K. (2011) Molecular dynamics simulation study of non-concatenated ring polymers in a melt. I. Statistics. *J. Chem. Phys.* **134**, 204904.

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

27. Barbieri,M., Chotalia,M., Fraser,J., Lavitas,L.M., Dostie,J., Pombo,A. and Nicodemi,M. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16173–16178.

28. Bohn,M. and Heermann,D.W. (2010) Diffusion-driven looping provides a consistent framework for chromatin organization. *PLoS One* **5**, e12218.

29. Jerabek,H. and Heermann,D.W. (2012) Expression-dependent folding of interphase chromatin. *PLoS One* **7**, e37525.

30. Jerabek,H. and Heermann,D.W. (2014) How chromatin looping and nuclear envelope attachment affect genome organization in eukaryotic cell nuclei. *Int. Rev. Cell. Mol. Biol.* **307**, 351381.

31. Stefano,M.D., Rosa,A., Belcastro,V., di Bernardo,D. and Micheletti,C. (2013) Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19. *PLoS Comput. Biol.* **9**, e1003019.

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

33. Canzio,D., Liao,M., Naber,N., Pate,E., Larson,A., Wu,S., Marina,D.B., Garcia,J.F., Madhani,H.D., Cooke,R. *et al.* (2013) A conformational switch in hp1 releases auto-inhibition to drive heterochromatin assembly. *Nature* **496**, 377–381.

34. Isono,K., Endo,T.A., Ku,M., Yamada,D., Suzuki,R., Sharif,J., Ishikura,T., Toyoda,T., Bernstein,B.E. and Koseki,H. (2013) Sam domain polymerization links subnuclear clustering of prc1 to gene silencing. *Dev. Cell* **26**, 565–577.

35. Timoshenko,E., Kuznetsov,Y.A. and Dawson,K. (1998) Conformational transitions of heteropolymers in dilute solutions. *Phys. Rev. E* **57**, 6801–6814.

36. Ramalho,T., Selig,M., Gerland,U. and Enßlin,T.A. (2013) Simulation of stochastic network dynamics via entropic matching. *Phys. Rev. E* **87**, 022719.

37. Doi,M. and Edwards,S. (1988) *The Theory of Polymer Dynamics*. Oxford University Press, Oxford.

38. van Kampen,N. (2001) *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam.

39. Press,W., Teukolsky,S., Vetterling,W. and Flannery,B. (2007) *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge.

40. Frenkel,D. and Smit,B. (2002) *Understanding Molecular Simulations: from Algorithm to Applications*. Academic Press, San Diego, CA.

41. Francis,N.J., Kingston,R.E. and Woodcock,C.L. (2004) Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574–1577.

42. Lo,S.M., Follmer,N.E., Lengsfeld,B.M., Madamba,E.V., Seong,S., Grau,D.J. and Francis,N.J. (2012) A bridging model for persistence of a polycomb group protein complex through DNA replication in vitro. *Mol. Cell* **46**, 784–796.

43. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master transcription factors and mediator establish superenhancers at key cell identity genes. *Cell* **153**, 307–319.

44. Kagey,M.H., Newman,J.J., Bilodeau,S., Zhan,Y., Orlando,D.A., van Berkum,N.L., Ebmeier,C.C., Goossens,J., Rahl,P.B., Levine,S.S. *et al.* (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435.

45. Timoshenko,E.G. and Kuznetsov,Y.A. (2002) Improved Gaussian self-consistent-method: Applications to homopolymers with different architectures in dilute solution. *J. Chem. Phys.* **117**, 5404–5414.

46. Fraser,P. and Bickmore,W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**, 413–417.

47. Mora,T. and Bialek,W. (2011) Are biological systems poised at criticality? *J. Stat. Phys.* **144**, 268–302.

48. Jost,D. (2014) Bifurcation in epigenetics: Implications in development, proliferation, and diseases. *Phys. Rev. E* **89**, 010701.

49. Noordermeer,D., Leleu,M., Splinter,E., Rougemont,J., Laat,W.D. and Duboule,D. (2011) The dynamic architecture of hox gene clusters. *Science* **334**, 222–225.

50. Naumova,N., Imakaev,M., Fudenberg,G., Zhan,Y., Lajoie,B.R., Mirny,L.A. and Dekker,J. (2013) Organization of the mitotic chromosome. *Science* **342**, 948–953.

51. Meuleman,W., Peric-Hupkes,D., Kind,J., Beaudry,J.B., Pagie,L., Kellis,M., Reinders,M., Wessels,L. and van Steensel,B. (2012) Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23**, 270280.

52. Vaquerizas,J.M., Suyama,R., Kind,J., Miura,K., Luscombe,N.M. and Akhtar,A. (2010) Nuclear pore proteins nup153 and megator define transcriptionally active regions in the Drosophila genome. *PLoS Genet.* **6**, e1000846.

53. Mukhopadhyay,S., Schedl,P., Studitsky,V.M. and Sengupta,A.M. (2011) Theoretical analysis of the role of chromatin interactions in long-range action of enhancers and insulators. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19919–19924.

54. Meluzzi,D. and Arya,G. (2012) Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Res.* **41**, 63–75.

55. Noordermeer,D. and Duboule,D. (2013) Chromatin architectures and hox gene collinearity. *Curr. Top. Dev. Biol.* **104**, 113–148.

56. Hawkins,R.D., Hon,G.C., Lee,L.K., Ngo,Q., Lister,R., Pelizzola,M., Edsall,L.E., Kuan,S., Luu,Y., Klugman,S., Antosiewicz-Bourget,J. *et al.* (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491.

57. Zhu,J., Adli,M., Zou,J.Y., Verstappen,G., Coyne,M., Zhang,X., Durham,T., Miri,M., Deshpande,V., Jager,P.L.D. *et al.* (2013) Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654.