

Thermodynamic Pathways to Genome Spatial Organization in the Cell Nucleus

Mario Nicodemi^{†‡*} and Antonella Prisco[§]

[†]Department of Physics and Complexity Science, University of Warwick, Coventry, United Kingdom; [‡]INFN Naples, Naples, Italy; and [§]CNR Institute of Genetics and Biophysics “Buzzati Traverso”, Naples, Italy

ABSTRACT The architecture of the eukaryotic genome is characterized by a high degree of spatial organization. Chromosomes occupy preferred territories correlated to their state of activity and, yet, displace their genes to interact with remote sites in complex patterns requiring the orchestration of a huge number of DNA loci and molecular regulators. Far from random, this organization serves crucial functional purposes, but its governing principles remain elusive. By computer simulations of a statistical mechanics model, we show how architectural patterns spontaneously arise from the physical interaction between soluble binding molecules and chromosomes via collective thermodynamics mechanisms. Chromosomes colocalize, loops and territories form, and find their relative positions as stable thermodynamic states. These are selected by thermodynamic switches, which are regulated by concentrations/affinity of soluble mediators and by number/location of their attachment sites along chromosomes. Our thermodynamic switch model of nuclear architecture, thus, explains on quantitative grounds how well-known cell strategies of upregulation of DNA binding proteins or modification of chromatin structure can dynamically shape the organization of the nucleus.

INTRODUCTION

Within cell nuclei, genomes have a complex organization in space spanning different scales. Chromosomes tend to form a set of distinct territories and, at a smaller level, are folded in higher-order structures, while a variety of physical intra- and interchromosomal interactions between specific DNA sequences take place (1–6). While structures can be formed by tethering specific DNA segments to scaffolding elements, such as the nuclear envelope, DNA-DNA contacts and chromatin loops are an ubiquitous organizational feature extending up to hundreds of kilobases, and relocating, for instance, genes to substantial distances outside of their territory. Intriguingly, relative positions of territories, as well as of DNA sequences within a territory, have a probabilistic nature dynamically changing with cell type and cell cycle phase. Yet, stable, nonrandom patterns are established, fundamental to genome regulation, as disruptions relate to serious diseases, most notably, cancer (1–6). Remarkably common features are shared in chromatin organization processes, but the underlying principles of their control in space and time are still largely mysterious (3).

Although there is evidence that far-apart DNA sequences, even on different chromosomes, can come together by interacting with molecular factors, the mechanisms whereby they do so and higher-order structures and territories that arise are still largely mysterious. One of the scenarios proposed to explain the establishment of contacts between DNA elements is the so-called random collision picture (see, e.g., (7)) whereby chromatin flexibility allows factors bound to one sequence to randomly contact factors bound to

surrounding chromatin. Although active mechanisms of directed motion have been described (see, e.g., (8)), diffusion-based mobility is indeed a prevailing mechanism that delivers molecular complexes to their specific nuclear targets (9). So, loops could be formed when a diffusing factor succeeds in bridging two chromosomal sites as a result of a random double encounter, whereby the molecule by chance encounters its first binding site and then, by chance, the second one. Yet how such loops persist beyond the initial random collision is totally unclear (7,10) and many questions remain open: how strong are the bonds required to hold in place whole chromosomal segments? How are stochastic encounters coordinated in space and time for a functional purpose by the cell? Can higher-order structures and territories spontaneously arise from them? Here, by use of a polymer physics model we propose a scenario to answer such questions.

Sequence-specific DNA-binding molecular factors have emerged as critical regulators of chromatin interactions in the nucleus (1–6) and some of them are encountered in a variety of cases, as for instance SATB1 (11), Ikaros (12), PcG (13), and CTCF Zn-finger proteins, the latter known to mediate also interchromosomal contacts (14–17). In some cases, a combination of factors is required to induce looping, as in the example of the erythroid transcription factor GATA-1 and its cofactors at the β -globin locus (18,19). Analogously, GATA-3 and STAT6 cooperation has been proposed to establish long-range chromatin interactions at the TH2 cytokine locus (20). Transcription factories themselves, i.e., local high concentrations of Pol II, have been proposed to act as hubs in the formation of loops and the colocalization of distant genes, even outside chromosome territories (5,21,22). In the last few years, protein-DNA

Submitted May 23, 2008, and accepted for publication December 9, 2008.

*Correspondence: mario.nicodemi@na.infn.it

Editor: Steven D. Schwartz.

© 2009 by the Biophysical Society
0006-3495/09/03/2168/10 \$2.00

doi: 10.1016/j.bpj.2008.12.3919

interactions that occur in vivo have been probed by innovative genome-wide techniques leading to the description of thousands of binding sites for DNA binding proteins (23), and systematic approaches to measuring their binding energy landscapes are being developed (24). DNA binding proteins typically exhibit a number of target loci, which can be found clustered in groups. Their DNA chemical affinities are in general found in the weak biochemical energy range, $E_X \sim 1 \div 15 \text{ } kT$ (24–28) (k is the Boltzmann constant and T room temperature). Although in most cases only qualitative information is available, details on binding energies and DNA locations have been clearly described for a number of examples (24–28). Initial works on bacteria have shown that DNA binding proteins can have hundreds of DNA sites with affinities in the range $2 \div 15 \text{ } kT$ (26). In yeast, more recently, the landscape of binding energies and loci has been explored by advanced computational biophysics methods: the distribution of their binding energies spans a range of $\sim 10 \text{ } kT$, and they can have hundreds of DNA binding sites across the genome as well (28). Similar ranges in binding energies have been found in higher eukaryotes, including mice and humans (23–25), where common examples exist of proteins with thousands of DNA target sequences.

DNA-DNA interactions mediated by molecular factors are being extensively mapped, revealing a complex network of intra- and interchromosomal interactions (29). Clusters of binding sites of SATB1 (11,30), and zinc finger class proteins CTCF (14,31,32), Ikaros (12), and GATA-1 (33) were found in a number of regions involved in DNA cross talk. An important example is the cluster of CTCF binding sites responsible for X chromosome pairing, at the onset of X inactivation, located at the *Xist/Tsix* locus where, in a few-kb-short sequence, a group of ~ 100 binding sites, each 20-b long, is found (17,31). Expansion of the nuclear volume leads to the disassembly of several nuclear compartments (34), which might suggest that specific concentrations of macromolecules are required for the self-assembly of nuclear structures. Loss of specific interchromosomal DNA-DNA contacts has been described after a marked reduction, for instance, of the amount of CTCF (15,17). Changes in the concentration of heterochromatin proteins, e.g., HP1 (35,36), are also known to affect the organization of genomic DNA (37).

The conformational properties of chromosomes have been investigated by using polymer models in the past (38–49). The chromatin fiber was modeled as a random walk in a confining geometry (38), and the possibility was considered to include giant loops, of $\sim 3 \text{ Mb}$, departing from its backbone to describe folding at different scales (RWGL model (39,40)). The multi-loop-subcompartment MLS model (41,42) aimed to represent rosette structures, with 120 kb loops, like those experimentally observed. To describe the radial arrangement of chromosome territories in human cell nuclei, a model was proposed (43) where each chromosome is approximated by a linear chain of spherical 1 Mbp-sized chromatin domains. Adjacent domains are linked together by an entropic spring

and by an effective excluded volume potential, while to maintain the compactness of chromosome territories a weak potential barrier around each chromosome chain was also included. Recently, the random loop polymer model (47) was introduced to explain, at the same time, the experimentally observed presence of loops of different scales and the leveling-off of the mean-square distance between two beads of the chain at genomic distances at $> 1\text{--}2 \text{ Mb}$. Several other chromatin features have been successfully explored by computer simulations, including nucleosome interactions (44), packing (45,46), and molecular assembly (48,49), providing a vivid description of the geometry and conformational properties of chromatin as observed in experiments.

Here, by investigation of a polymer physics model inspired by the above biological scenario, we discuss how architectural patterns spontaneously arise from the interaction of soluble-binding molecules and chromosomes. Our model shows that thermodynamics dictates pathways to complex pattern formation: loops, colocalization of distant sequences, chromosomal domains, structures, and territories spontaneously organize as stable thermodynamic states when specific threshold values in molecule concentrations or their affinity to DNA sites are exceeded. By regulation of expression levels and modification of DNA targets, the cell can, thus, act on thermodynamic switches (50,51) to reliably control its genome organization in space and time.

THEORETICAL MODEL

To describe a system made of a chromosome and its binding molecules, we consider an established model of polymer physics (52,53): the chromosome polymer is modeled as a Brownian self-avoiding walk (SAW) of n nonoverlapping beads, and soluble molecules as Brownian particles having a concentration, c (Fig. 1). A fraction, f , of polymer sites can bind the diffusing molecules, with a chemical affinity E_X in the weak biochemical range (see Methods for details). Here, for sake of simplicity, binding sites are uniformly interspersed with nonbinding regions along the chain. Each molecular factor can simultaneously bind many a site on the polymer, a feature that reflects the presence of multiple DNA binding domains in a number of regulatory proteins (e.g., CTCF). Mediating molecules with only one DNA binding site, that are able to interact with each other, could be also considered; since a group of linked molecules can be represented, in the model, as just one mediator, the picture is unchanged. The equilibrium thermodynamic properties of such a system were determined by extensive Monte Carlo simulations (53,54).

METHODS

In our Monte Carlo computer simulations (53) molecules and polymers diffuse in a cubic lattice having a linear size L , and its spacing, d_0 , sets the space unit. For computational purposes, we mostly consider lattices of linear size $L = 32$, although we tested our results up to $L = 128$. SAW

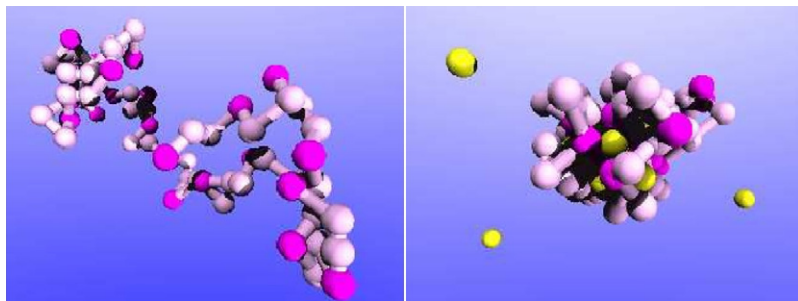


FIGURE 1 Model. The figure shows two representative snapshots from our three-dimensional computer simulations. In the left panel, a self-avoiding walk (SAW) polymer is shown, as it floats randomly within the assigned volume without forming stable loops. In the right panel the volume also contains a concentration $c = 0.04\%$ of Brownian molecules (yellow) having an affinity $E_X = 4$ kT for a fraction $f = 1/3$ of the polymers beads (shown in a darker shade). As molecules can bind more than one polymer site, loops can be formed. However, they are stable, and confine the polymer in a closed territory (as in the case shown here), only if c is above a threshold value (see Fig. 2). The SAW chains shown here comprise $n = 64$ beads.

polymer beads have a diameter, d_0 , and each bead in a chain is on a next or nearest next-neighboring site of its predecessor. Molecules (of size d_0) are also subject to Brownian motion. When neighboring a binding site of a polymer, molecules interact with it via an effective energy, E_X . According to the studied case (see Results), up to six distinct sites (i.e., the nearest neighbors in a cubic lattice) on the same chromosome, or alternatively two sites on different chromosomes, can be bound at the same time.

Our schematic model is a coarse-grained description of a real polymer and, since by now we mostly focus on the description of a general conceptual framework, beads only represent generic binding sites (they could be a binding locus, the bases of specific binding sequences, etc.). In cases where detailed data on binding sequences and regulator chemistry is available, such information could be easily taken into account in the model to produce specific quantitative predictions. The role of interactions with, e.g., the nuclear membrane could be also included, but to make the message simpler, we decided not to discuss such an aspect here.

To obtain thermodynamic equilibrium configurations, the Metropolis Monte Carlo method was applied. Chromosome polymers are initially equilibrated in a random self-avoiding configuration obtained, in absence of binding molecules, by random displacements of single beads under the constraint that each bead in the chain is on a next or nearest next-neighboring site of its predecessor. Then molecules are inserted at random empty positions in the lattice to attain a given concentration. In the ensuing Metropolis Monte Carlo procedure, a sequence of states is generated by a Markov process (53), whereby a new position for a particle/bead is stochastically selected according to a specific transition matrix satisfying the principle of detailed balance, which, in turn, guarantees the convergence in probability of the sampled states to Boltzmann thermodynamic equilibrium distribution. The transition probability for a particle/bead to diffuse to a neighboring empty site is proportional to the Arrhenius factor $r_0 \exp(-\Delta E/kT)$, where ΔE is the energy barrier in the move, k the Boltzmann constant, and T the temperature (53). The lattice has periodic boundary conditions to reduce boundary effects.

In a Monte Carlo lattice sweep, every particle and bead in the system, randomly selected, is updated on average once. Our simulations run for up to 10^9 Monte Carlo lattice sweeps as the number of decorrelation steps from an initial configuration can be as large as 10^5 . The achievement of stationarity was monitored by checking the dynamics of different quantities, such as the system gyration radius, the distance between two polymers, the system energy, and the number of particles attached to polymers. Once equilibrium is reached for all these quantities, thermodynamic averages are calculated by considering only configurations having a distance larger than the decorrelation length. Finally, averages are also performed over up to 2048 runs from different initial configurations. Confidence intervals are calculated as squared deviations around these averages (53); they are indicated in our figures by the size of the used symbols.

Our code has two core routines, well described in Binder and Heermann (53): the lattice gas spin-exchange Metropolis routine for particle displacement, and the SAW routine. Several means were considered to avoid algorithmic errors, as those suggested by Binder and Heermann (53). Each different routine in the code was tested independently. For example, the routine generating the evolution of the SAW chain was tested by checking

the behavior of the calculated average gyration radius, R_g , against the chain length, n , and the power law $R_g \sim n^\nu$ with an exponent $\nu \sim 0.6$, well established in the literature (52,53), was recovered. Another internal test was to show that other geometric quantities, such as the chain end-to-end distance, did scale in the same way as R_g .

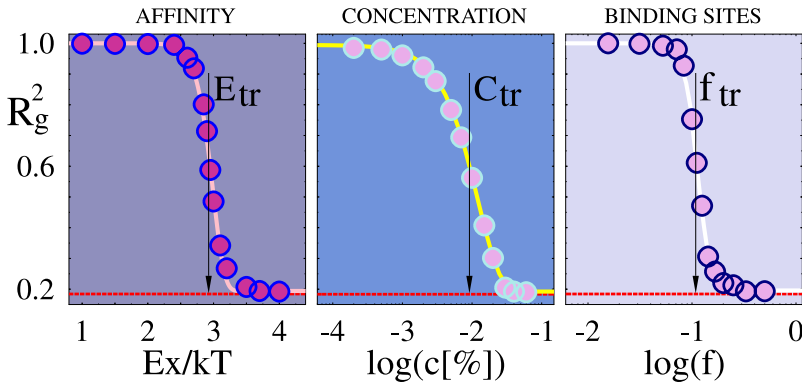
Real chromosomes differ in size (i.e., n) and arrangement of their binding sites. Such differences affect their specific behaviors, but the general picture we aim to depict here is not altered by changes to the selected values of these parameters (e.g., n and L). To make computation time feasible, we mostly use $n = 64$, but we tried n as large as 128. The robustness of our model is well established in polymer physics (52,53), and to check the effects of finite size scaling, we explored changes of the polymer chain length in the range $n \in \{16, 32, 64, 128\}$ (see Results).

RESULTS

Intrachromosomal interactions, loop, and territory formation

We first discuss how a chromosome can fold up in loops within a territory with a specific spatial conformation by interacting with soluble molecules, and how the process can be controlled by the cell (see Fig. 1). The folding state of the polymer is illustrated by its squared radius of gyration, R_g^2 (52): $R_g^2 = 1/(2n^2N) \sum_{i,j=1}^n (\mathbf{r}_i - \mathbf{r}_j)^2$, where \mathbf{r}_i is the position of bead $i \in \{1, \dots, n\}$, and N a normalization constant (here N equals the average squared gyration radius of a randomly floating SAW chain of size n). The value R_g represents the radius of a minimal sphere enclosing the polymer: it attains a maximum when the polymer is loose and randomly folded, and a minimum when loops enclose it in a compact lump.

In presence of a given concentration of molecules, loops could be created by chance when a particle bridges a couple of chromosomal sites having a nonzero affinity, E_X . Fig. 2 (left panel) shows, indeed, that R_g^2 attains a small plateau value when E_X is large enough (say above the inflection point, E_{tr} , of the curve $R_g^2(E_X)$): bridges are thermodynamically favored and the polymer takes a compact looped territorial conformation, as seen in a typical snapshot from computer simulations depicted in Fig. 1 (right panel). The system behavior, however, switches for $E_X < E_{tr}$, since R_g^2 keeps its maximal value corresponding to a fully open polymer floating in space (see Fig. 1, left panel) and no stable loops are formed. The folding level also depends on factors



mer forms a looped territory. In the central panel, R_g^2 is shown as a function of c , for a given E_X and f (here $E_X = 4$ kT, $f = 1/3$). In addition, in this case a threshold effect is observed ($c_{tr} \approx 0.01\%$), although a broader crossover region exists where the level of folding can be tuned. The right panel shows the sharp threshold of R_g^2 as a function of f ($f_{tr} \approx 0.1$, here $c = 0.04\%$, $E_X = 4$ kT), illustrating that only in presence of multiple sites (i.e., above f_{tr}) the polymer can be folded in loops. In all the above cases, loops are thermodynamically stable only above the threshold values, as a consequence a phase transition occurring in the system. By tuning affinities/concentrations, the cell can act, thus, on a thermodynamic switch to form and release loops and territories.

such as concentration of molecules, number, and location of DNA binding sites (see below).

The above results have an intuitive basis: if E_X is small, the half-life of a randomly formed bridge is small and polymer segments on average float away; the higher E_X , the higher the number of bound molecules and, thus, of bridges which reinforce each other and stabilize the conformation, as multiple bonds should be simultaneously broken to release a loop. Our physics model reveals, in particular, that a precise threshold marks the switch between the two regimes; E_{tr} corresponds to a thermodynamic phase transition (55), as discussed later on. This picture illustrates on quantitative grounds how chromatin modifications, such as DNA methylation or post-translational modifications of DNA binding proteins (well described cell strategies to change genomic architecture), can result in dramatic, switchlike, effects.

In a different thermodynamic pathway to loop formation, the cell can regulate the concentration, c , of binding molecules. The plot of $R_g^2(c)$ (Fig. 2, central panel) shows how c affects the compaction state of the polymer. When c is below the threshold, c_{tr} , $R_g^2(c)$ has a value corresponding to random folding, while $>c_{tr}$, it decreases toward its looped state value. A broad crossover region is found at $\sim c_{tr}$, revealing that R_g^2 , which can be envisioned in our example as the radius of the territory, can be tuned across a range of values. Therefore, the regulation of a DNA binding protein concentration (a typical event in cellular behavior) can act as another switch to reliable assembling of genomic architectures.

Finally, we find (Fig. 2, right panel) that a minimal threshold in the number of polymer binding sites (or in their fraction f) is required for stable looping/territory formation. Conceptually, the case of a polymer with a low number of binding sites is equivalent to the case of a polymer with many binding sites in the presence of a limiting concentration of mediators (Fig. 3). The function $R_g^2(f)$ indicates that a thermodynamic switch to DNA compaction resides

in the potential to obliterate/restore a fraction of sites via chromatin modifications that abolish binding of the relevant regulatory molecule. Intriguingly, the presence of a

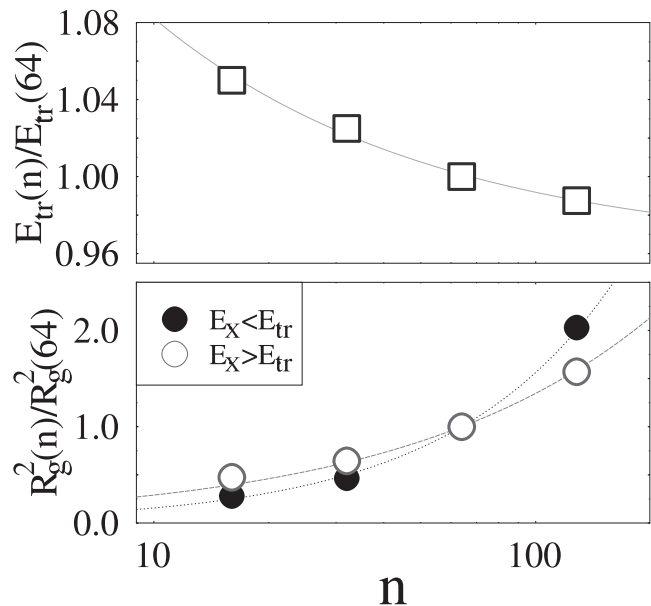


FIGURE 3 Scaling properties. (Lower panel) The average gyration radius, R_g , relative to polymer model considered into the left panel of Fig. 2, is plotted as a function of the polymer chain length n . The picture shows the ratio $R_g^2(n)/R_g^2(64)$ (since $n = 64$ is the reference case dealt with in the rest of the article) for $n = 16, 32, 64, 128$. In the phase where the polymer is open, i.e., for $E_X = 1$ kT $< E_{tr}$ (see left panel of Fig. 2), the average gyration radius, R_g (solid circles), scales with n as a power law $R_g \sim n^\nu$ with an exponent $\nu \sim 0.6$ (52,53) (superimposed fit, dashed line). In the looped phase, i.e., for $E_X = 4$ kT $> E_{tr}$, R_g (open circles) scales as $n^{1/3}$ (superimposed fit, long dashed line), showing that the polymer is lumped in a compact conformation. (Upper panel) The threshold energy, E_{tr} , relative to the left panel of Fig. 2, is a function of the polymer chain length n . Here we plot the ratio $E_{tr}(n)/E_{tr}(64)$ (where $E_{tr}(64) \approx 3$ kT). The superimposed fit is $E_{tr}(n) = E_{tr}^\infty + A/n^B$, where $E_{tr}(n)$ is the threshold energy for a chain of size n , $E_{tr}^\infty \sim 0.96E_{tr}(64)$, the extrapolated value for an infinitely long system, $A \sim 0.47E_{tr}(64)$, and $B \sim 0.5$ a fitting coefficient and exponent.

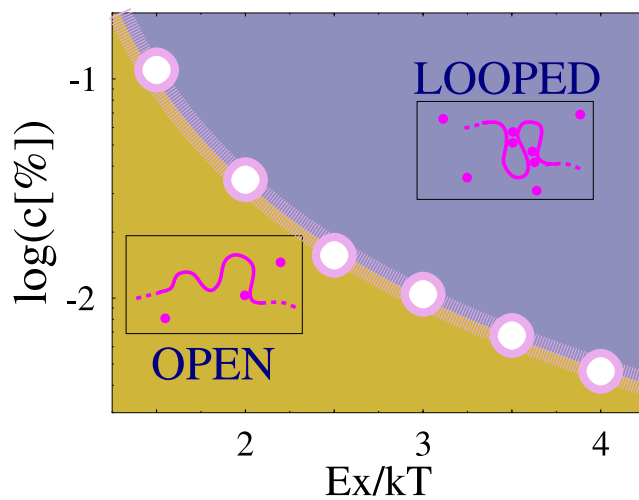


FIGURE 4 Phase diagram. The state of the polymer/chromosome (see Fig. 1) at thermodynamic equilibrium is summarized by this phase diagram in a range of values of weak biochemical affinities, E_X , and concentration, c , of its binding molecules (here $f = 1/3$). When E_X and c are below the transition line, $E_{tr}(c)$ (open circles), the polymer is open (as sketched in the inset) and no stable loops can be formed. Above threshold, instead, the system enters the region where the polymer is folded and looped on itself.

thermodynamic threshold in f could relate to the experimental observation that multiple binding sites for mediators have been found at chromosome interaction loci and looping points (e.g., CTCF-mediated interactions). Importantly, in our model we find that the threshold value, f_{th} , is a strongly decreasing function of the binding energy, E_X . This can be expected as, for an above threshold mediator concentration, c , the overall binding energy linking two polymer strands is $\sim fE_X$; so an increase of E_X would correspond to an inversely proportional reduction of f_{tr} .

The above-described thermodynamic switches define a robust regulatory mechanism as seen in the phase diagram of Fig. 4, reporting the equilibrium state of the chromosome (open versus looped) in a wide range of E_X and c values (for a given f). In particular, Fig. 4 shows that the threshold value $E_{tr}(c)$ (dashed line) required for loop assembly decreases as c increases and can be as weak as a hydrogen bond. In the cell, the possibility to drive looping by use of sites with even low binding energy for their soluble ligands could be important to prevent polymers from getting stuck in topologically unacceptable entanglements or ectopic associations, since each single low energy bond can be easily broken for adjustments.

The threshold values in the (c, E_X, f) space (see Fig. 4), related in polymer physics to the chain θ -point (52), correspond to a phase transition occurring in the system when one of two competing thermodynamics mechanisms prevails: entropy, S , which favors loose random folding; or energy, E , which increases when bonds between molecules and DNA sites are established by loop formation. The system spontaneously tends (as it is finite-sized (55)) to select the state where its free energy, $F(c, E_X, f) = E - TS$, is mini-

mized. More precisely, the chromosome conformation has a specific stochastic distribution (having a width which can be very narrow) following from Boltzmann thermodynamics weights (55).

Scaling behavior of the model

As molecule binding regions on cross-talk loci of real chromosomes have variable sizes, n , we explored the scaling behavior of our system by varying the polymer chain length in the range $n \in \{16, 32, 64, 128\}$, for the above value of the containing box size L . The reference case considered previously, and in the rest of the article, has $n = 64$, which is comparable to values included in similar studies (41–43).

We investigate, in particular, how the average gyration radius, R_g , and the threshold energy, E_{tr} , depend on n . For a matter of clarity, we refer to the case discussed in the left panel of Fig. 2, but similar features are found for the other cases presented in our article. We, thus, consider a system with $c = 0.04\%$ and $f = 1/3$, and discuss first the case where $E_X = 1$ kT, i.e., the phase where the polymer is open (see Fig. 2, left panel). Under these circumstances, as shown in the lower panel of Fig. 3, R_g scales with n as a power law, $R_g \sim n^\nu$, with an exponent $\nu \sim 0.6$, which is in agreement with the random SAW scaling laws (52,53). Conversely, for $E_X = 4$ kT, i.e., in the looped phase, R_g scales as $n^{1/3}$ (see lower panel of Fig. 3), showing that the polymer is lumped in a compact conformation ($1/3$ is the inverse of the Euclidean dimension of the system). The threshold energy E_{tr} has also a comparatively simple behavior with n and appears to saturate to a finite value for large n . For instance, the threshold energy defined in the left panel of Fig. 2 (where $E_{tr}(64) \sim 3$ kT) can be well fitted by a power law in n (see upper panel of Fig. 3): $E_{tr}(n) = E_{tr}^\infty + A/n^B$, where $E_{tr}(n)$ is the value for a chain of size n , E_{tr}^∞ the fitting value for an infinitely long chain, A and B a fitting coefficient and exponent (we find $E_{tr}^\infty \simeq 0.96E_{tr}(64)$, $A \sim 0.47E_{tr}(64)$ and $B \sim 0.5$).

Similar properties are found for the other quantities discussed in this article. These checks outline the robustness of the picture discussed above and also support the idea that it is not an artifact of discretization, as a system in the continuum limit, i.e., on a finely divided lattice, should have an analogous behavior.

Interchromosomal segment interactions

The mechanisms that drive other layers of spatial organization, including the colocalization of DNA sequences belonging to different chromosomes (56) and the relative positioning of chromosomal territories (1–6), can be shown to be very similar to those inducing stable loop formation within a single chromosome. Concentration/affinity acts in these cases as a thermodynamic switch for segment colocalization and for chromosome positioning in a map.

To such an aim, in an extension of the model described above, we now investigate the thermodynamic state of

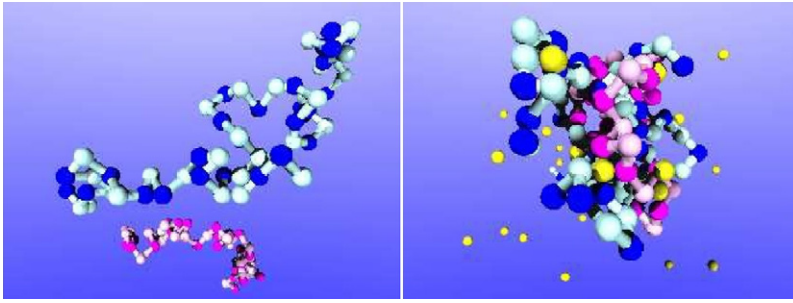


FIGURE 5 Two snapshots are shown from computer simulations of our two polymer model. In the left panel, the polymers float independently within the assigned volume. In the right panel, the volume also includes a concentration, $c = 0.3\%$, of molecules (yellow particles), which can bind simultaneously each polymer once at any of their specific loci (darker sites, here in a fraction $f = 1/2$ with affinity $E_X = 4 kT$). When c is above a threshold value (see Fig. 6), as in the case shown, thermodynamically stable bridges can be formed between the polymers, which spontaneously tend to pair parts of, or all of, their chains.

two SAW chains (representing either two distal sequences on the same chromosome or sequences on distinct chromosomes) with a fraction f of binding sites (periodically placed) for a concentration, c , of molecules having an affinity, E_X , to both of them (Fig. 5); for simplicity, each molecule can bind once either polymer. The relative polymer positioning is given by their squared distance: $d^2 = 1/(2n^2D) \sum_{i,j=1}^n (\mathbf{r}_i^{(1)} - \mathbf{r}_j^{(2)})^2$, where $\mathbf{r}_i^{(1)}$ (respectively, $\mathbf{r}_i^{(2)}$) is the position of bead i in chromosome 1 (respectively, 2), and Δ a normalization constant (here Δ is equal to the average square distance of two independent random SAW chains). The average value of d^2 is maximal when polymers float independently (i.e., $d^2 = 1$ in our normalization) and decreases drastically when all or parts of the chains become colocalized.

Regulation of E_X can induce formation or release of stable physical contact between the polymers. Fig. 6 shows that when E_X is below a threshold, E_{tr} , their equilibrium distance, d^2 , has the same value found for two noninteracting Brownian SAW chains (i.e., $d^2 = 1$). This is the random phase where chromosomes move independently. By thermodynamics mechanisms an effective attraction between the polymers is, instead, established when $E_X > E_{tr}$: physical contact is stable and d^2 drastically decreased, as the system enters the colocalization phase. The equilibrium distance is a function of c as well (see Fig. 6): when c is below a threshold value,

c_{tr} , a random distance is found between chromosomes (i.e., $d^2 = 1$). Colocalization is spontaneously attained, instead, when c increases, as $d^2(c)$ approaches a plateau with a much smaller value. Finally, for a given c and E_X , colocalization can be achieved only if the number of binding sites along the polymers is above a sharp threshold value, as shown in Fig. 6 where $d^2(f)$ is plotted.

As with loop architecture within a chromosome territory, the average distance of chromosome pairs can be controlled via thermodynamics mechanisms. The spatial association is attained when a phase transition line is crossed, corresponding to the point where entropy loss due to chain pairing is compensated by energy gain as both polymers are bound: the lower the E_X , the higher the concentration, c , required.

Assembly of chromosome territorial maps

Within the above picture, the relative positioning of chromosomal loci and territories can be understood by similar arguments. As an example, we considered (see inset in Fig. 7) the case with three SAW chains ($n = 64$) having each a fraction, f , of binding sites ($f = 1/2$, $E_X = 4 kT$): the sites on polymer 1 and 2 interact with a molecular factor (concentration c_{12}) which can bind once either chain; polymer 2 and 3 bind a different molecular factor of concentration c_{23} (for definiteness, we only discuss the case where $c_{12} = c_{23} = c$). To illustrate the important effects of physical interference between

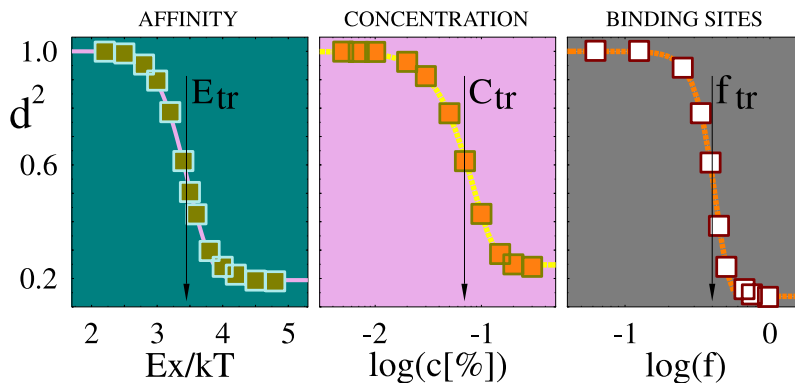


FIGURE 6 Thermodynamic switches for interchromosomal contacts. The equilibrium average distance, d^2 , of the two polymer model pictured in Fig. 5, is a function of the affinity, E_X , of their binding sites for diffusing molecules; of the concentration, c , of molecules; and of the fraction, f , of polymer binding sites. In the left panel, d^2 is plotted as a function of E_X (here $c = 0.3\%$, $f = 1/2$). When E_X is smaller than a threshold, $E_{tr} \sim 3.5 kT$, d^2 is maximal ($d^2 = 1$ in our normalization) and the polymers float independently one from the other. For $E_X > E_{tr}$, d^2 drastically decreases, as the polymers are spontaneously colocalized. In the central panel, d^2 is shown as a function of c (here $E_X = 4 kT$, $f = 1/2$) and a threshold appears as well ($c_{tr} \sim 0.07\%$), surrounded by a crossover region. In the right panel, the sharp threshold of d^2 as a function of f is

shown ($f_{tr} \sim 0.4$, here $E_X = 4 kT$, $c = 0.3\%$): only multiple binding sites, above f_{tr} , can achieve polymer colocalization. The mechanism driving polymer colocalization is an effective reciprocal attraction of thermodynamic origin, related to a phase transition: below threshold, molecules bridging by chance the polymers do not succeed in holding them in place; above threshold, bridges are thermodynamically stabilized. Molecular mediators act, then, as a thermodynamic switch to spontaneous formation and release of polymer stable contacts.

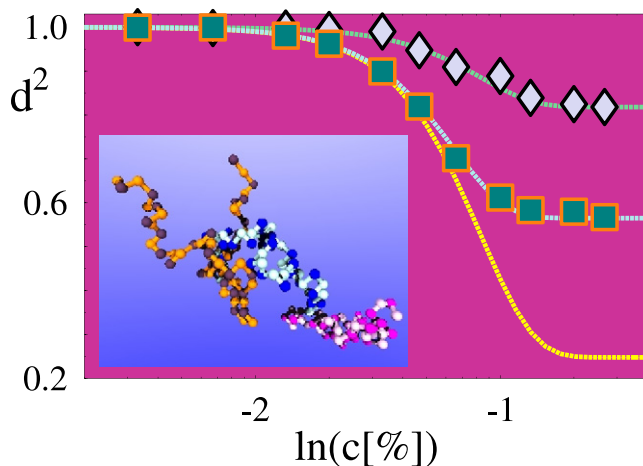


FIGURE 7 Territorial maps. The relative positions of three polymers can be regulated by the concentration of specific molecular factors. (*Inset*) A configuration is shown from our computer simulations of a three polymer model. A specific molecular factor can bind polymers 1 (pink) and 2 (blue), while a different factor binds polymers 2 and 3 (orange). Both molecular factors have here a concentration $c = 0.13\%$ ($E_X = 4 kT$, $f = 1/2$), but they are not shown for clarity. (*Main panel*) The average distance between polymers 1 and 2, d_{12}^2 (squares), decreases as a function of c (the distance between 2 and 3 equals d_{23}^2 , and is not shown). As an indirect effect of the attraction within pairs 1-2 and 2-3, the distance between 1 and 3, d_{13}^2 (diamond), decreases as well, remaining, however, above d_{12}^2 . The three polymers, thus, tend to form a triangle with two short equal edges (corresponding to d_{12} and d_{23}) and a longer edge (i.e., d_{13}). In general, by tuning c , E_X , and f , a variety of configurational patterns can be spontaneously attained. Notably, since polymers 1 and 3 compete for bridging the sites of polymer 2, they physically interfere and d_{12}^2 is larger than in the case of an isolated couple (yellow lower line, from Fig. 6). A proper spatial organization of chromosomes in territories and within territories could also help in minimizing physical interference and entanglement.

chromosomes, in this model all molecular factors compete for the same sites on polymer 2. For the built-in symmetry, polymer couples 1-2 and 2-3 behave similarly and have, on average, equal relative distances $d_{12}^2 = d_{23}^2$ as a function of c (see Fig. 7). Yet, since polymer 1 and 3 physically interfere when bridging with 2, in a competition for its binding sites, their distance is larger than the one found in the case with only two polymers under similar conditions (i.e., same c , E_X , f and system size). The distance between 1 and 3, d_{13} , is in turn larger than $d_{12} = d_{23}$ because there is not a direct interaction. The three chromosomes, thus, spontaneously find their position to form a (isosceles) triangle having sides of predefined length (d_{12} , d_{23} , d_{13}).

Different patterns of relative positions can be attained by tuning the concentration/affinity switches, as the system architecture self-organize via thermodynamics pathways, funneling the interaction between sets of DNA binding sites and matching molecular mediators. When the number and length of chromosomal segments increase, the dynamics of the system to equilibrium can be slowed down by physical hindrance. Thus rises the speculation that the spatial organization of chromosomes in distinct territories and

within territories (along with other mechanisms, e.g., the action of topoisomerases) may also serve the purpose of a faster and better control of their interaction and function, by reducing undesired entanglements.

DISCUSSION

Within the cell nucleus, in a striking example of self-organization, an astonishing number and diversity of DNA loci and molecular mediators are spatially orchestrated to form a complex and functional architecture involving regulatory cross talk between distant sites. We propose a simple conceptual framework, a thermodynamic switch model of nuclear architecture, to understand some of its general features, namely (4):1), how a chromosome can fold up into a territory and how its looping is dynamically controlled by binding molecules;2), how chromosomes interact and establish their relative positioning;3), what the regulatory principles are; and 4), the origin of the stochastic character of territorial maps.

Our model consists of a system of SAW polymers interacting with soluble molecular mediators. By use of statistical mechanics, we have shown that thermodynamics dictates pathways to complex pattern formation, via mechanisms such as thermodynamic switches (Fig. 8). This supports, on quantitative grounds, the idea that a variety of intra- and interchromosome interactions can be traced back to similar mechanisms. Looping and compaction, remote sequence interactions, and territorial-segregated configurations correspond to thermodynamic states selected by appropriate values of concentrations/affinity of soluble mediators and by number and location of their attachment sites along chromosomes. After proper concentrations/affinities are set, the organization proceeds spontaneously with no energetic costs as the resources required, e.g., to rearrange even whole chromosomes, are provided by the surrounding thermal bath. Our picture explains, thus, how well-described cell strategies of upregulation of DNA binding proteins or modification of chromatin structure can shape the genomic architecture and produce DNA colocalization and territories according to thermodynamically driven nonrandom patterns.

Testable quantitative predictions are shown on the biological effects of alterations of genomic DNA sequences (such as deletions, insertions, chemical changes, etc.) and of their molecular mediators (concentrations, binding energies, etc.). In particular, the model highlights the fact that, at above-threshold values of concentration, the interaction with low affinity molecular factors may be sufficient to drive the compaction of chromosomes into territories, and shows that the interaction of chromosomes with soluble mediators has the potential to impart a probabilistic relative arrangement to chromosomes. Our analysis reveals that molecular factors that act as bridges between two chromosomes may not only have the effect of pulling those close to each other, but may also displace noninteracting chromosomes, so that

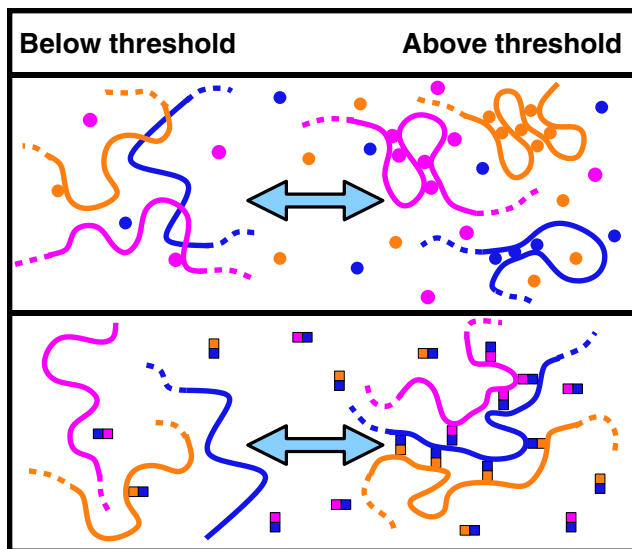


FIGURE 8 Schematic illustration of thermodynamic switches and their effects at different levels of system organization. (*Top panel*) The assembly of intrachromosomal contacts and loops is thermodynamically possible only when the concentration/affinity of binders (circles) exceeds precise threshold values. At that point, previously randomly and independently diffusing molecules and chromosomes spontaneously generate an organized pattern, in a process reversible by downregulation of the switch. Specific conformations can be attained by site specificity of a set of molecular mediators. (*Bottom panel*) Similar threshold and self-organization mechanisms act for establishing contact between remote loci and, at a higher scale, relative positions of territories. A variety of patterns, encoded in the location of a number of binding sites along chromosomes, can be precisely selected via thermodynamics effects by a combinatorial use of a set of molecular mediators (rectangles).

these are farther away from each other than the random distance. This result is thought-provoking in the light of experimental data (56) showing that disruption of transcription can lead either to an increase or to a decrease of chromosome intermingling among specific couples of chromosomes, depending on which two chromosomes you look at. Allele-specific, parent-of-origin specific, and expression-specific DNA-DNA interactions have also been described (15,16,57–59). In this context, our analysis could explain how imprinting and other allele-specific protein-DNA interactions may have the capacity to address homologous chromosomes to two different regions of the territory map.

A rough estimate of threshold molecular concentrations in real nuclei can be made from our predicted concentration values: here c is the number of molecules per lattice site, so the number of molecules per unit volume is c/d_0^3 , where d_0 is the linear lattice spacing constant. The molar concentration ρ is obtained by dividing by the Avogadro number N_A . Note that threshold concentrations depend on the binding energy E_X (see, e.g., Fig. 4). For sake of definiteness, however, we can consider the case with $E_X \sim 2 kT$ (see Fig. 4), where threshold concentrations are $\sim c = 0.1\% \div 0.01\%$. Under the rough assumption that d_0 is a couple of orders-of-magnitude smaller than the nucleus diameter (i.e., $d_0 \sim 10$ nm), a threshold molar concentration would be $\rho \sim 0.1 \div 1 \mu\text{mole/}$

liter, which is consistent with typical experimental values of nuclear protein concentrations (60,61). Such estimation is very rough, but may help to further bridge this study with biological investigations.

Starting from experimental results showing that chromatin fiber at large genomic distances, $>1\text{--}2$ Mb, exhibits a leveling-off of the mean square distance between two DNA sites, a Gaussian random loop polymer model was recently proposed (47). To explain these observations, the model proposed that a given number P of couples of distant beads, randomly selected along the chain, are bound by an harmonic potential of amplitude κ . The model investigated the mean distance between sites and the size of loops, and showed that the presence of random loops on all length scales explains the leveling-off of the mean-square distance. That model is close in spirit to this work where cross interactions of a fraction, f , of DNA sites are mediated by the binding of molecular factors and by the formation of bridges of energy E_X . In our case, the number of interacting site couples also depends on the concentration of mediators, c , and their interaction is short-ranged. Nonetheless, the number of interacting site in our model would correspond, in the notation of Bohn et al. (47), to a P , which is a (nontrivial) function of c . These considerations can illustrate the agreement between the discovery in Bohn et al. (47) of the leveling-off of the mean-square distance and our finding, for instance, that for c above threshold, the polymer gyration radius does not attain the (self-avoiding) random walk value but saturates to much smaller values.

In real cells, passive and active regulatory mechanisms can cooperate, adding further layers of complexity (1–6), while the list of molecules mediating chromatin organization is likely to include dedicated structural proteins, RNAs and, e.g., the transcription, replication, or repair machinery (21,62,63). In our picture, specificity of interactions is obtained by specific molecular mediators binding to specific loci, while other general molecules could help the process. In the arrangement of specific binding sites along chromosomes and scaffolding elements, a variety of spatial patterns can be encoded (43) on an evolutionary timescale. Within a cell, patterns could be then dynamically selected by the combinatorial use of a set of mediators via the ineluctable, yet probabilistic, laws of thermodynamics (45).

The authors declare no competing interests.

Work supported by MIUR-FIRB grant No. RBNE01S29H, Network No. MRTN-CT-2003-504712.

REFERENCES

1. Cremer, T., M. Cremer, S. Dietzel, S. Muller, I. Solovei, et al. 2006. Chromosome territories—a functional nuclear landscape. *Curr. Opin. Cell Biol.* 18:307–316.
2. Lanctôt, C., T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer. 2007. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8:104–115.

3. Misteli, T. 2007. Beyond the sequence: cellular organization of genome function. *Cell*. 128:787–800.
4. Meaburn, K., and T. Misteli. 2007. Chromosome territories. *Nature*. 445:379–381.
5. Fraser, P., and W. Bickmore. 2007. Nuclear organization of the genome and the potential for gene regulation. *Nature*. 447:413–417.
6. Akhtar, A., and S. Gasser. 2007. The nuclear envelope and transcriptional control. *Nat. Rev. Genet.* 8:507–517.
7. de Laat, W., and F. Grosvel. 2003. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.* 11:447–459.
8. Chuang, C., A. Carpenter, B. Fuchsova, T. Johnson, P. de Lanerolle, et al. 2006. Long-range directional movement of an interphase chromosome site. *Curr. Biol.* 16:825–831.
9. Misteli, T. 2001. Protein dynamics: implications for nuclear architecture and gene expression. *Science*. 291:843–847.
10. de Laat, W. 2007. Long-range DNA contacts: romance in the nucleus? *Curr. Opin. Cell Biol.* 19:317–320.
11. Galande, S., P. Purbey, D. Notani, and P. Kumar. 2007. The third dimension of gene regulation: organization of dynamic chromatin loop-scale by SATB1. *Curr. Opin. Genet. Dev.* 17:408–414.
12. Brown, K., S. Guest, S. Smale, K. Hahm, M. Merkenschlager, et al. 1997. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*. 91:845–854.
13. Bantignies, F., C. Grimaud, S. Lavrov, M. Gabut, and G. Cavalli. 2003. Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. *Genes Dev.* 17:2406–2420.
14. Kurukuti, S., V. Tiwari, G. Tavoosidana, E. Pugacheva, A. Murrell, et al. 2006. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. USA*. 103:10684–10689.
15. Ling, J., T. Li, J. Hu, T. Vu, H. Chen, et al. 2006. CTCF mediates inter-chromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*. 312:269–272.
16. Zhao, Z., G. Tavoosidana, M. Sjlander, A. Göndär, P. Mariano, et al. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38:1341–1347.
17. Xu, N., M. Donohoe, S. Silva, and J. Lee. 2007. Evidence that homologous X-chromosome pairing requires transcription of CTCF protein. *Nat. Genet.* 39:1390–1396.
18. Drissen, R., R. Palstra, N. Gillemans, E. Splinter, F. Grosveld, et al. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* 18:2485–2490.
19. Vakoc, C., D. Letting, N. Gheldof, T. Sawado, M. Bender, et al. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell*. 17:453–462.
20. Spilianakis, C., and R. Flavell. 2005. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* 5:1017–1027.
21. Osborne, C., L. Chakalova, K. Brown, D. Carter, A. Horton, et al. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36:1065–1071.
22. Marenduzzo, D., I. Faro-Trindade, and P. Cook. 2007. What are the molecular ties that maintain genomic loops? *Trends Genet.* 23:126–133.
23. Massie, C., and I. Mills. 2008. ChIPping away at gene regulation. *EMBO Rep. Rev.* 9:337–343.
24. Maerkl, S., and S. Guake. 2007. A system approach to measuring binding energy landscape of transcription factors. *Science*. 315:233–237.
25. Morozov, A., J. Havranek, D. Baker, and E. Siggia. 2005. Protein-DNA binding specificity predictions with structural models. *Nucleic Acids Res.* 33:5781–5798.
26. Gerland, U., J. Moroz, and T. Hwa. 2002. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc. Natl. Acad. Sci. USA*. 99:12015–12020.
27. Lässig, M. 2007. From biophysics to evolutionary genetics: Statistical aspects of gene regulation. *BMC Bioinform.* DOI:10.1186/1471-2105-8-S6-S7, Published: 27 September 2007.
28. Berg, J. 2008. Dynamics of gene expression and the regulatory inference problem. *Europhys. Lett.* 82:28010.
29. Simonis, M., J. Kooren, and W. de Laat. 2007. An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods*. 4: 895–901.
30. Purbey, P., S. Singh, P. Kumar, S. Mehta, K. Ganesh, et al. 2008. domain-mediated dimerization and homeodomain-directed specificity are required for high-affinity DNA binding by SATB1. *Nucleic Acids Res.* 36:2107–2122.
31. Donohoe, M., L. Zhang, N. Xu, Y. Shi, and J. Lee. 2007. Identification of a CTCF cofactor, Yy1, for the X chromosome binary Switch. *Mol. Cell*. 25:43–56.
32. Scialdone, A., and M. Nicodemi. 2008. Mechanics and dynamics of X-chromosome pairing at X inactivation. *PLoS Comp. Biol.* 5, e10002444.
33. Jing, H., C. Vakoc, L. Ying, S. Mandat, H. Wang, et al. 2008. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol. Cell*. 29:232–242.
34. Hancock, R. 2004. A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus. *J. Struct. Biol.* 146:281–290.
35. Hediger, F., and S. Gasser. 2006. Heterochromatin protein 1: don't judge the book by its cover! *Curr. Opin. Genet. Dev.* 16:143–150.
36. Kwon, S., and J. Workman. 2008. The heterochromatin protein 1 (HP1) family: put away a bias toward HP1. *Mol. Cells*. 26:217–227.
37. Locke, J., M. Kotarski, and K. Tartof. 1988. Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect. *Genetics*. 120:181–198.
38. Hahnfeldt, P., J. Hearst, D. Brenner, R. Sachs, and L. Hlatky. 1993. Polymer models for interphase chromosomes. *Proc. Natl. Acad. Sci. USA*. 90:7854–7858.
39. Yokota, H., G. van den Engh, J. Hearst, R. Sachs, and B. Trask. 1995. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the Human G0/G1 interphase nucleus. *J. Cell Biol.* 130:1239–1249.
40. Sachs, R., G. Engh, B. Trask, H. Yokota, and J. Hearst. 1995. A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl. Acad. Sci. USA*. 92:2710–2714.
41. Munkel, C., and J. Langowski. 1998. Chromosome structure predicted by a polymer model. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*. 57:5888–5896.
42. Munkel, C., R. Eils, S. Dietzel, D. Zink, C. Mehning, et al. 1999. Compartmentalization of interphase chromosomes observed in simulation and experiment. *J. Mol. Biol.* 285:1053–1065.
43. Kreth, G., J. Finsterle, J. von Hase, M. Cremer, and C. Cremer. 2004. Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code. *Biophys. J.* 86:2803–2812.
44. Mergell, B., R. Everaers, and H. Schiessel. 2004. Nucleosome interactions in chromatin: fiber stiffening and hairpin formation. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 70:011915.
45. Marenduzzo, D., C. Micheletti, and P. Cook. 2006. Entropy-driven genome organization. *Biophys. J.* 90:3712–3721.
46. Bon, M., D. Marenduzzo, and P. Cook. 2006. Modeling a self-avoiding chromatin loop: relation to the packing problem, action-at-a-distance, and nuclear context. *Structure*. 14:197–204.
47. Bohn, M., D. Heermann, and R. van Driel. 2007. Random loop model for long polymers. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76: 051805.
48. Nicodemi, M., and A. Prisco. 2007. A symmetry breaking model for X chromosome inactivation. *Phys. Rev. Lett.* 98:108104.

49. Nicodemi, M., and A. Prisco. 2007. Self-assembly and DNA binding of the blocking factor in X chromosome inactivation. *PLoS Comp. Biol.* 3:2135–2142.
50. Nicodemi, M., B. Panning, and A. Prisco. 2008. A thermodynamic switch for chromosome colocalization. *Genetics*. 179:717–721.
51. Nicodemi, M., B. Panning, and A. Prisco. 2008. Colocalization transition of homologous chromosomes at meiosis. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 77:061913.
52. Doi, M., and S. Edwards. 1984. *The Theory of Polymer Dynamics*. Clarendon Press, Oxford, UK.
53. Binder, K., and D. Heermann. 2002. *Monte Carlo Simulation in Statistical Physics*. Springer-Verlag, Berlin, Heidelberg, and New York.
54. Binder, K. 1997. Applications of Monte Carlo methods to statistical physics. *Rep. Prog. Phys.* 60:487.
55. Stanley, H. 1971. *Introduction to Phase Transitions and Critical Phenomena*. Clarendon Press, Oxford, UK.
56. Branco, M., and A. Pombo. 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* 4:e138.
57. Maynard, N., J. Chen, R. Stuart, J. Fan, and B. Ren. 2008. Genome-wide mapping of allele-specific protein-DNA interactions in human cells. *Nat. Methods*. 5:307–309.
58. Lomvardas, S., G. Barnea, D. Pisapia, M. Mendelsohn, J. Kirkland, et al. 2006. Interchromosomal interactions and olfactory receptor choice. *Cell*. 126:403–413.
59. Takizawa, T., P. Gudla, L. Guo, S. Lockett, and T. Misteli. 2008. Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes Dev.* 22:489–498.
60. Handwerger, K., J. Cordero, and J. Gall. 2005. Cajal bodies, nucleoli, and speckles in the *Xenopus oocyte* nucleus have a low-density, sponge-like structure. *MBC*. 16:202–211.
61. Hancock, R. 2007. Packing of the polynucleosome chain in interphase chromosomes: evidence for a contribution of crowding and entropic forces. *SCDB*. 18:668–675.
62. Cook, P. 2002. Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* 32:347–352.
63. Osborne, C., L. Chakalova, K. Brown, D. Carter, A. Horton, et al. 2005. Replication and transcription: shaping the landscape of the genome. *Nat. Rev. Genet.* 6:669–677.