

## Research Article

# DNA Loci Cross-Talk through Thermodynamics

**Antonio Scialdone<sup>1,2</sup> and Mario Nicodemi<sup>2,3</sup>**

<sup>1</sup> *Dipartimento di Scienze Fisiche, Università di Napoli “Federico II”, 80126 Napoli, Italy*

<sup>2</sup> *INFN Sezione di Napoli, 80126 Napoli, Italy*

<sup>3</sup> *Department of Physics and Complexity Science, University of Warwick, Coventry CV4 7AL, UK*

Correspondence should be addressed to Antonio Scialdone, antonio.scialdone@na.infn.it

Received 26 March 2009; Revised 5 June 2009; Accepted 8 July 2009

Recommended by Serdar Kuyucak

The recognition and pairing of specific DNA loci, though crucial for a plenty of important cellular processes, are produced by still mysterious physical mechanisms. We propose the first quantitative model from Statistical Mechanics, able to clarify the interaction allowing such “DNA cross-talk” events. Soluble molecules, which bind some DNA recognition sequences, produce an effective attraction between distant DNA loci; if their affinity, their concentration, and the relative DNA binding sites number exceed given thresholds, DNA colocalization occurs as a result of a thermodynamic phase transition. In this paper, after a concise report on some of the most recent experimental results, we introduce our model and carry out a detailed “in silico” analysis of it, by means of Monte Carlo simulations. Our studies, while rationalize several experimental observations, result in very interesting and testable predictions.

Copyright © 2009 A. Scialdone and M. Nicodemi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

In the complex mechanisms of the functional regulation of genome, the spatial distribution of DNA loci plays a key role. Actually, the genome nuclear architecture deeply affects genes transcription status and can determine the success or the failure of several cellular processes [1–8]. A vital, nontrivial genome arrangement is the result of a number of interactions between distant DNA loci which can pair and/or associate with nuclear scaffolding elements (e.g., membrane, matrix, etc.).

A wealth of examples can be cited: from the homologous chromosomes recognition and pairing, necessary, for example, during meiosis [9–11], to the clustering of genes around specific nuclear structures, such as the “speckles,” which can determine their active/silenced status [4].

Longstanding unresolved issues are the mechanisms whereby distant DNA loci are able to “cross-talk” in order to achieve the right space configuration, and how the cell is able to organize in space and time such events [12].

Here, we set in a Statistical Mechanics framework the most recent discoveries about the molecular bases of such “DNA cross-talk” phenomena. As a result, we show how it

is possible to build a general, thermodynamic-based model, which answers the above questions.

We refer to a well-established biological picture, whereby a DNA locus needs the presence of soluble binding molecules to target another DNA segment. Starting from these experimental evidences, we analyzed a Physics model describing the interaction between a couple of polymers and a concentration  $c$  of binding molecules in a lattice. We showed that these two ingredients—say, a set of soluble molecules able to bind a number of DNA sites—are sufficient to enable a “DNA cross-talk,” which eventually results in the colocalization of the DNA loci. By means of extensive Monte Carlo computer simulations, we carried out the first quantitative analysis of this kind of phenomena. We dissected in fine details how the DNA cross-talk is influenced by changes of some parameters, such as the binding molecules concentration, the DNA binding sites number, and their relative affinity. We observed a threshold effect: the DNA loci colocalization is possible provided that the above mentioned parameters pass specific threshold values. Only in this case the DNA loci stop their Brownian diffusion and, as a result of a thermodynamic phase transition, pair off.

While this rationalizes the experimentally proved importance for DNA cross-talks of the molecular factors concentration and of the length of the relative DNA attaching regions (see [13] and below), it suggests some possible regulatory mechanisms by means of well-described cellular strategies, like the upregulation of a DNA binding molecules or the modification of chromatin structure. Moreover, our analysis of the dynamics of DNA colocalization, which is currently under investigation in several wet experiments, provides some interesting, testable results.

In the following, before introducing our model in more details, we briefly report on some of the many experiments, which point out the special role played by precise DNA segments and specific DNA binding molecules for the “DNA cross-talks”.

*Protein-Mediated Chromatin Loops.* A variety of examples are known of chromatin loops, with fundamental regulatory roles, which are the result of molecule-mediated interactions between different DNA loci. One of these occurs during *T* cells maturation. A naive *T* cell can differentiate into a TH1 or a TH2 cell depending on which locus among the *Ifng* on chromosome 10 or the *Th2* on chromosome 11 becomes activated [14, 15]. The cell maturation is accompanied by a change in the chromatin architecture within the above mentioned loci. In particular, on TH2 cell activation, the chromatin at *Th2* locus adopts a transcriptionally active structure consisting of a series of small loops. These loops are realized thanks to SATB1 protein, which is rapidly expressed upon the cell maturation and binds 9 sites within *Th2* locus [14]. Another example is the occurrence of two chromatin loops at *Kit* gene, which are achieved through the contacts between different DNA loci mediated by GATA proteins [16]. The loop able to activate *Kit* is produced by the binding of GATA-2 protein to the loci-situated +5 kB and -114 kB with respect to the *Kit* promoter. Instead, GATA-1 intervenes to mediate the interaction between the loci +58 kB and +5 kB in order to form the loop which downregulates *Kit*. So, the control of the active/downregulated status of *Kit* gene, essential for erythropoiesis (the production of red blood cells), is thought to be accomplished through the regulation of the relative expression level of GATA-1 and GATA-2 proteins [16].

*X Chromosomes Dynamic Arrangement during X-Chromosomes Inactivation.* An even more striking event, which entails the organization of two whole chromosomes, happens during X-Chromosomes Inactivation (XCI). XCI is one of the most mysterious aspects in current mammalian X biology. It is the phenomenon occurring in the female mammalian cells, which leads to the silencing of one of its X chromosomes randomly chosen, to equalize the dosage of X products with respect to males, having only one X [17–20]. The X chromosomes “cross-talk” and their spatial arrangement is fundamental for the success of XCI. At the beginning of process, the XCI regulatory sequences on the X’s couple, called *Xic* (X-Inactivation Center), have to be juxtaposed [21, 22]. The interaction allowing this colocalization is

mediated by a protein-RNA bridge which includes CTCF protein, able to bind a cluster of about 50 sites in the *Tsix/Xite* sequence within *Xic* [22]. The random choice of the X to be inactivated follows and, then, while the active X moves to the nuclear membrane, the inactive X associates with the nucleolus, to maintain its silenced status [23].

## 2. Results and Discussion

*2.1. The Model.* On the basis of the experimental evidences, we built a schematic model for DNA cross-talk, including a couple of DNA segments endowed with a number of binding sites (BSs) for a set of molecular factors (MFs) (see Figure 1).

For sake of simplicity, DNA segments as well as MFs move in a lattice whose spacing is  $d_0$ , of the order of magnitude of the BS length (see Section 4). The lattice dimensions are  $L_x = 2L$ ,  $L_y = L$  and  $L_z = L$  in units of  $d_0$ .

DNA segments are described via a standard model of polymer physics [24]. They are formed by  $L$  nonoverlapping beads, each included in a single lattice site, which randomly diffuse under the constraint that two proximal beads must be on next or nearest next neighboring sites (nonbreaking constraint). In each DNA locus, a number  $n_0$  of beads are BSs.

MFs are represented by Brownianly diffusing beads, and, to a first approximation, only a single MF at a time can be present in a lattice site. When an MF and a polymer BS are in neighboring sites, they interact via an effective energy  $E$ ; MFs can bind a couple of BSs from both DNA segments at the same time (in resemblance of several known mediating proteins, e.g., CTCF; see Figure 1). The energy function  $\mathcal{H}$  of the system is the following:

$$\mathcal{H} = -E \sum_{\langle \vec{r}, \vec{r}' \rangle} n_B(\vec{r}) n_M(\vec{r}'), \quad (1)$$

where  $n_B(\vec{r}) = 0, 1$  is the occupation variable of BSs; it is 1 if the lattice site at position  $\vec{r}$  is occupied by a BS; otherwise it is 0 (a site cannot be occupied by more than one kind of particle). Similarly,  $n_M(\vec{r}') = 0, 1$  is the occupation variable of MFs. The sum runs on all nearest neighboring lattice sites  $\vec{r}$  and  $\vec{r}'$ . We studied equilibrium and dynamic features of this system by means of Monte Carlo (MC) computer simulations.

Our schematic model does not include all the complex DNA regulatory phenomena taking place within the cell nucleus. However, the cross-talking mechanism we envisaged here is general, since it grounds on thermodynamic bases (see next sections), and for this reason it is not affected by the complexity of the model. Thus, while the approximations we use do not have any influence on cross-talking efficiency, they allow us to explore a wide range of parameters combinations, without the drawback of computational unfeasibility.

As for the parameter values, we chose weak biochemical energy values for  $E$  ( $\sim 0/5kT$ ), typical for DNA-protein interactions [25–28]. The MF concentration,  $c$ , (expressed as the ratio between the MFs number and the lattice sites

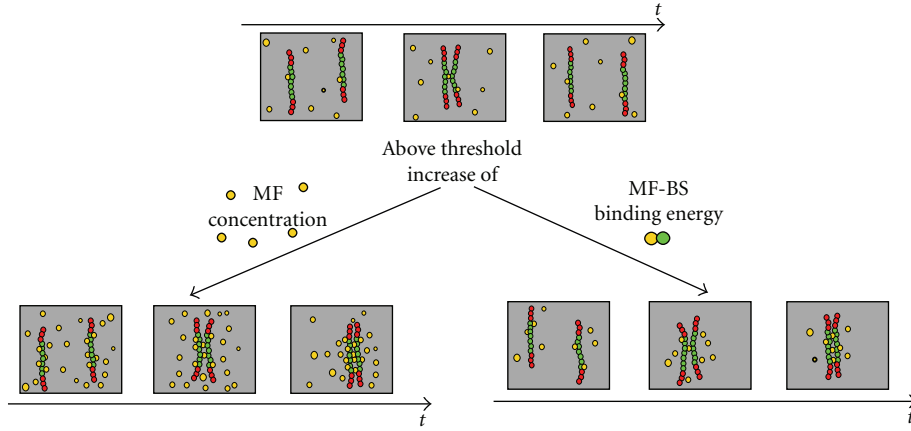


FIGURE 1: A molecular factors-mediated interaction produces DNA colocalization. The pictures schematically show how DNA segments at equilibrium colocalize, when the concentration of Molecular Factors (MFs, yellow beads) or their affinity for the Binding Sites (BSs, green beads) exceeds a threshold value. Actually, when this happens, thermodynamically stable “MF bridges” are built between the DNA loci.

number), spanned the range 0.01/1%, which corresponds to typical nuclear protein concentration (see Section 4). The binding sites number  $n_0$  is 24 (i.e., of the order of magnitude of the known case of CTCF binding sites in *Tsix/Xite* [29]); however, we also varied it to verify the BSs deletion effects (see Figure 4). The lattice we considered has  $L = 32$  and periodic boundary conditions (see Section 4 for further details).

Time is given in units of Monte Carlo lattice sweeps [30]; details about the conversion from MC time to real time are given in the Section 4.

All the following results are averaged over up to 500 runs.

**2.2. The Dynamics of Pairing.** When an MF concentration  $c$  is present within the lattice and the polymers have  $n_0$  BSs each, an effective attraction is induced between the couple of polymers, whose intensity depends on the value of the MF-BS interaction energy  $E$ .

This is shown by the normalized mean square distance between the polymers, defined as

$$d^2(t) = \frac{1}{n_0} \sum_{z=1}^{n_0} \frac{\langle r^2(z, t) \rangle}{L^2}, \quad (2)$$

that is, the square distance between two beads at the same height  $z$ ,  $r^2(z, t)$ , averaged over the  $n_0$  BSs and over different realizations (as the symbols  $\langle \dots \rangle$  indicate).  $L^2$  is used for normalization.

In Figure 2 the plot of  $d^2(t)$  with  $c = 0.2\%$  and  $n_0 = 24$  for two values of the energy  $E$  is illustrated. At  $t = 0$  the MFs were randomly placed within the lattice sites, and the polymers were in their straight vertical configuration, at a distance  $L$  from each other (so  $d^2(0) = 100\%$ ). It is evident that the increase of  $E$  from 1.4 kT to 2.2 kT determines very different equilibrium states for the system: while at lower energy,  $d^2$  saturates at the value expected from a couple of independent random walk polymers ( $d^2 \sim 40\%$ ); when  $E = 2.2$  kT,  $d^2$  eventually goes down to  $\sim 0$ , thus revealing

that the effective attraction is now high enough to make DNA loci pair off.

$d^2(t)$  shows an initial linear behaviour, determined by the random diffusion of the polymers when they are far from each other, and a final exponential saturation to a plateau which is dependent, as we saw, on the values of  $E$ . A good fit function, which includes both the above mentioned time regimes, is the following (superimposed fits in Figure 2(a)):

$$d^2(t) = d^2(\infty) + \left[ d^2(0) - d^2(\infty) + \frac{at}{1+bt} \right] e^{-t/\tau}, \quad (3)$$

where  $d^2(\infty)$  is the final equilibrium value, while  $a$ ,  $b$ , and  $\tau$  are fit parameters which depend on  $E$ ,  $c$ , and  $n_0$ .

**2.3. An MF-Mediated Interaction.** To get more insight into the dynamics of this “cross-talk mechanism” and into the key role of the molecular factors, we monitored the time behaviour of the average fraction  $F$  of the MFs attached to a single DNA locus. The role of MFs as mediators of the DNA loci interaction clearly emerges in Figure 2(b), where the plot of  $F(t)$ , for the same values of the parameters used above, is shown together with  $d^2(t)$  fit functions.

For both the energies  $E$ ,  $F(t)$  starts from the same value, due to the initial random distribution of the MFs in the lattice sites. When  $E = 1.4$  kT,  $F(t)$  simply saturates in a time scale which is about 4 orders of magnitude smaller than that of  $d^2(t)$ ; actually, MFs have a size much smaller than polymers, so their dynamics is much faster. If the energy is increased to  $E = 2.2$  kT and the DNA loci pair off ( $d^2$  saturates at  $\sim 0$ ), a more interesting time evolution is found. In fact, now, two time regimes are easily distinguished: during the first  $\sim 0.5$  minute,  $F(t)$  has a behaviour similar to that in the previous case, except for the higher plateau due to the higher MF-BS affinity. This first time regime corresponds to the initial MF binding to the DNA loci which, still out of the action range of the MF-induced effective potential, independently diffuse. Yet, just when  $d^2$  begins to exponentially fall down to  $\sim 0$ ,  $F(t)$  rises to a second plateau:

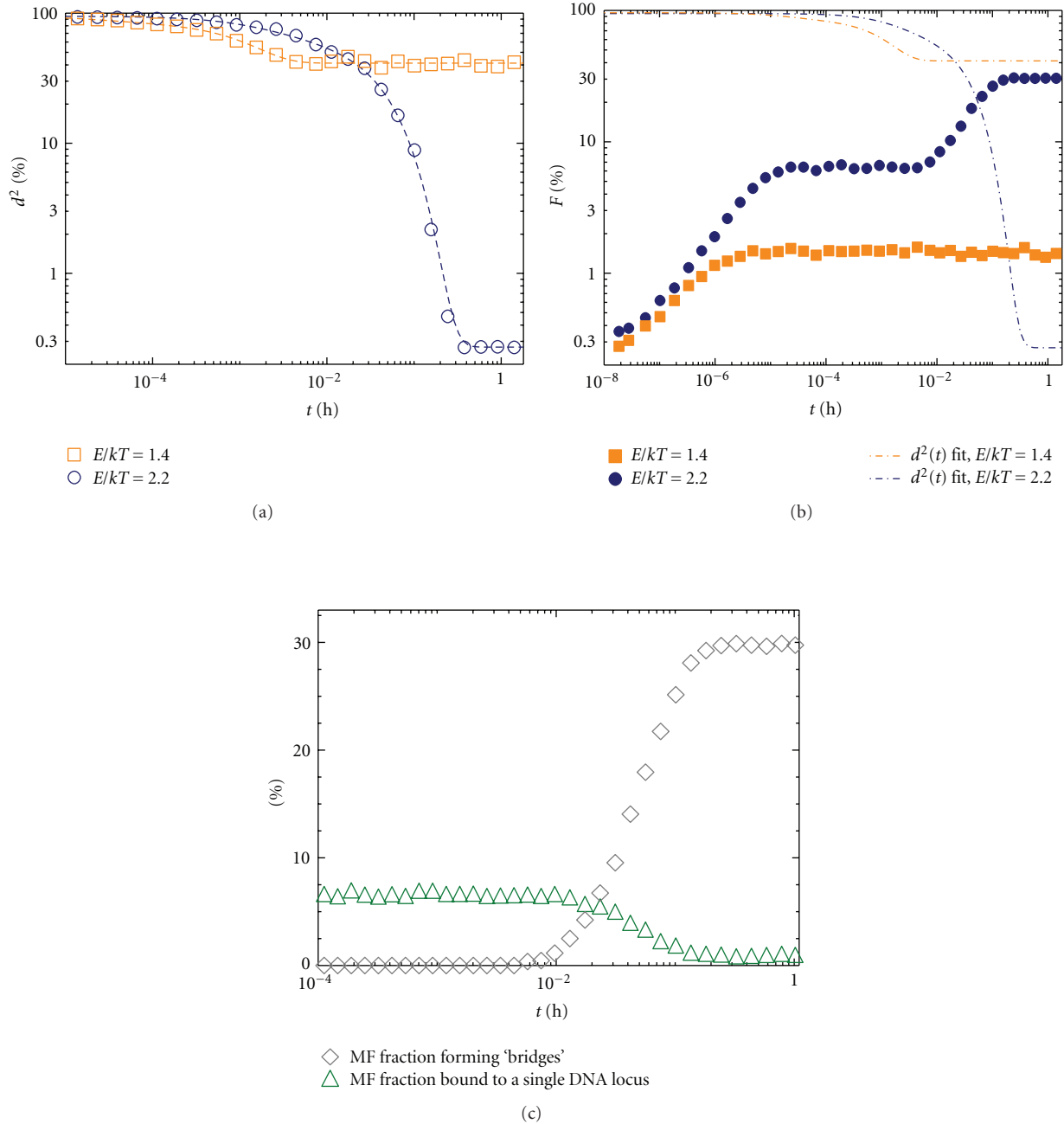


FIGURE 2: A couple of DNA loci colocalizes by increasing the MF-BS interaction energy  $E$ . (a) The normalized mean square distance  $d^2$  is shown as function of  $t$  for two values of the interaction energy  $E$  between the Molecular Factors (MFs) and the Binding Sites (BSs). DNA loci start at a distance of the order of the containing volume size ( $d^2 \sim 100\%$ ). Then, for  $E/kT = 1.4$  (orange square markers),  $d^2$  saturates at the Brownian value ( $\sim 40\%$ ), while, at  $E/kT = 2.2$  (blue circle markers), an effective attraction appears, so strong to produce DNA colocalization at equilibrium ( $d^2 \sim 0\%$ ). (b) Together with the fit function for  $d^2(t)$  (dashed lines), the mean fraction of MFs attached to a single DNA locus,  $F(t)$ , is plotted.  $F(0)$  does not depend on  $E$ , since MFs were initially distributed at random in the lattice. Subsequently,  $F(t)$  reaches a first plateau, which is due to the MF binding to the single DNA locus and is, obviously, higher with larger  $E$ . Yet, while at lower energy ( $E/kT = 1.4$ , orange square markers) this first plateau is also the thermodynamic equilibrium value for  $F(t)$ ; if  $E/kT = 2.2$  (blue circle markers),  $F(t)$  rises to a second plateau, just when  $d^2(t)$  exponentially decreases to  $\sim 0$ . Actually, at this value of  $E$ , at equilibrium, some additional MFs bind both DNA loci to form thermodynamic stable “bridges”(see Figure 1), which keep them together. This is illustrated in (c), where the MF fraction bind to both DNA loci (gray diamonds) and to a single DNA locus (green triangles) is plotted as function of  $t$  for  $E/kT = 2.2$  (the sum of these two curves gives  $F$  for  $E = 2.2kT$ , see (b)). Here an MF concentration  $c = 0.2\%$  is present, with  $n_0 = 24$  BSs within each DNA locus.

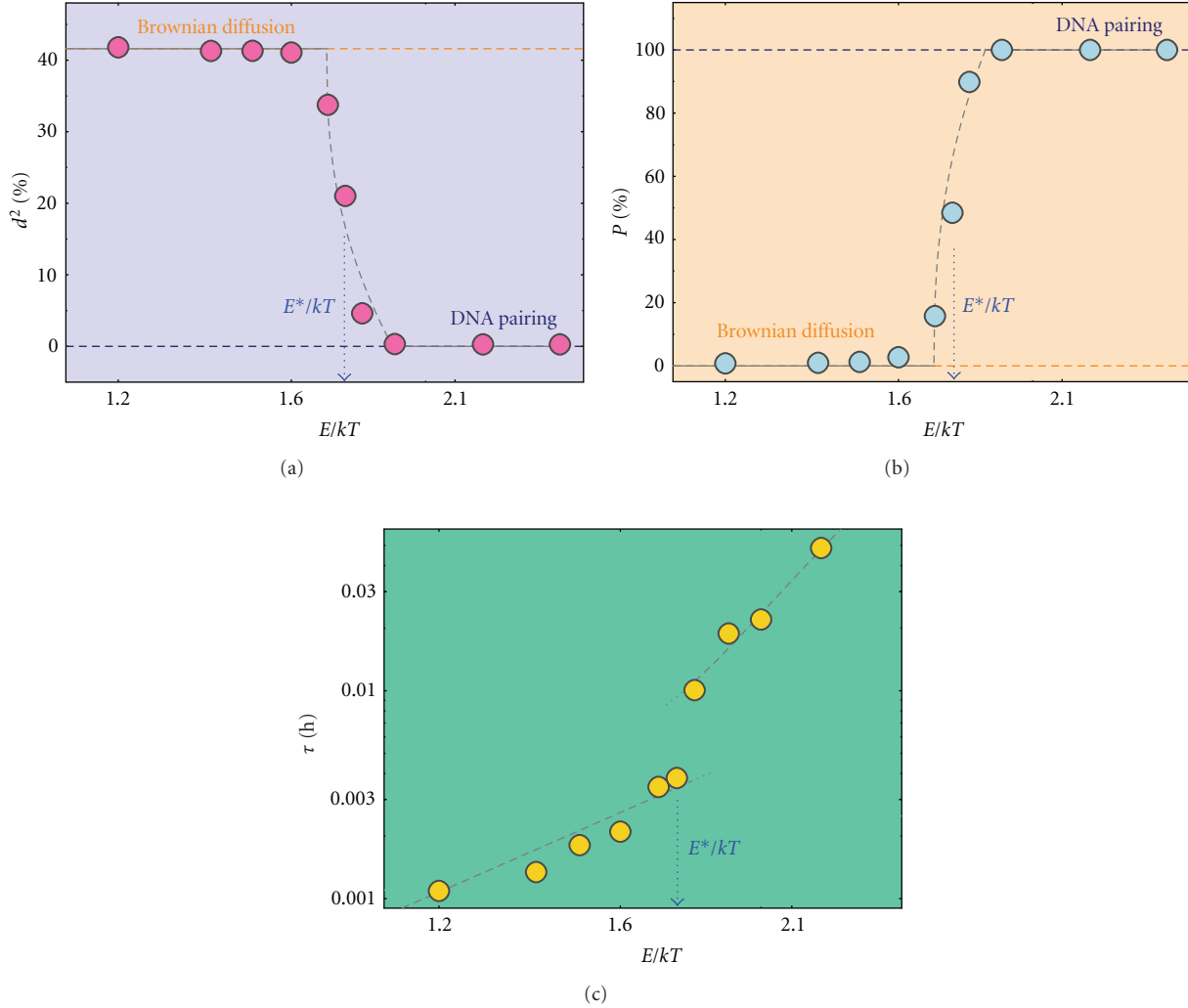


FIGURE 3: DNA loci colocalization is the result of a thermodynamic phase transition. We show this through the plots of the normalized mean square distance ( $d^2$ , (a)) and of the fraction of paired DNA loci ( $P$ , (b)) at equilibrium, as functions of  $E$ , the interaction energy between the molecular factors (MFs) and the DNA binding sites (BSs). As  $E$  increases,  $d^2$  and  $P$ , from values typical of Brownian diffusion ( $d^2 \sim 40\%$  and  $P \sim 0\%$ , orange horizontal lines), rapidly saturate to values which signal DNA loci full colocalization ( $d^2 \sim 0\%$ ,  $P \sim 100\%$ , blue horizontal lines), after a threshold value  $E^*/kT = 1.75$  (defined by the criterion  $P(E^*) = 50\%$ ). Superimposed fits for both  $d^2(E)$  and  $P(E)$  are power-law functions (gray dashed lines). In (c) the characteristic time scale  $\tau$  needed to reach thermodynamic equilibrium is illustrated, as function of  $E$ .  $\tau(E)$  turned out to be an increasing function, as the stronger MF-BS bonds make the DNA loci diffusion more difficult. A “jump” is observed at  $E^*$ , when the phase transition takes place. The superimposed fits are power-law functions. We took an MF concentration  $c = 0.2\%$  and a DNA BS number  $n_0 = 24$ .

as soon as DNA loci, during their random diffusion, get closer, some MFs start to stably bind both of them, and, as a result,  $F(t)$  increases. This is evidenced in Figure 2(c), showing that quite all the bound MFs actually form “bridges” between the DNA loci after  $\sim 1$  minute. These MFs “bridges” keep together the two DNA loci, so that pairing is finally produced (see Figure 1).

The  $F(t)$  analysis, while allows an immediate distinction between two dynamical regimes, makes the MF mediating role evident: we saw that colocalization takes place as a result of MF bridges realization, which, however, occurs at thermodynamic equilibrium only for certain values of the parameters. To better understand this issue, we now carry out

an analytic, though approximate calculation for the pairing probability  $P$ .

**2.4. An Approximate Formula for the Pairing Probability.** For sake of simplicity, we considered the two DNA segments close to each other in their straight configuration. If a random distribution of a concentration  $c$  of MFs is presumed, a number  $cn_0$  of them will be present, on average, between the  $n_0$  BSs of the DNA loci. For each of these MFs two states are possible: they can form a “bridge” or not. The energy equal to  $-2E$  corresponds to the former case, while the energy is 0 in the latter. Thus, according to the canonical distribution,

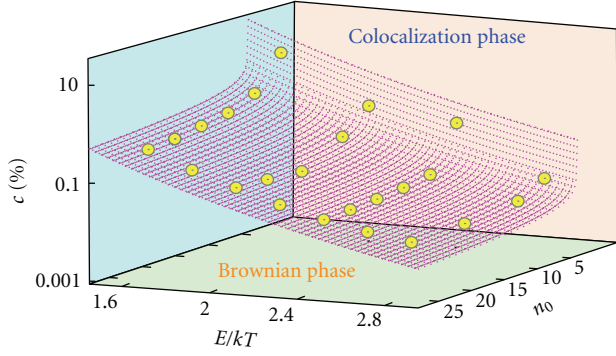


FIGURE 4: The DNA colocalization happens when specific threshold values in  $(c, E, n_0)$  are passed. A 3D phase diagram in the space of the molecular factors (MFs) concentration  $c$ , the binding sites (BSs) number  $n_0$ , and the MF-BS interaction energy  $E$  is shown. When the purple surface is crossed, a transition from a “Brownian Phase” (DNA loci diffuse as random walk polymers) to a “Colocalization Phase” (colocalized DNA loci) occurs. The yellow circles mark the measured transition points (at which the 50% of DNA loci couples are paired); the purple surface was obtained by a power law fit ( $cE^\gamma n_0^\delta = \text{const.}$ ,  $\gamma = 4$ ,  $\delta = 1.1$ ).

$p = \exp(2E/kT)/(1 + \exp(2E/kT))$  and  $q = 1/(1 + \exp(2E/kT))$  are, respectively, the probability that a single MF bridge is formed or not.

We can define the pairing probability  $P$  as the probability that at least one MF bridge is built between the DNA loci. Since  $cn_0$  is the total number of the possible MF bridges, the following formula holds:

$$P = 1 - (q)^{cn_0} = 1 - \frac{1}{[1 + \exp(2E/kT)]^{cn_0}} \quad (4)$$

according to which  $P$  increases with  $(c, E, n_0)$ .

While this calculation, in its simplicity, helps understanding how the MFs presence induces an effective attraction at thermodynamic equilibrium which is stronger with higher  $(c, E, n_0)$ , in the following section we seek for the exact thermodynamics of the system by means of MC simulations.

**2.5. The Thermodynamic Phase Transition.** What emerged from MC simulations is that, even if the intensity of the effective potential always increases with  $(c, E, n_0)$ , the system reacts to a change in one of these parameters only if a threshold value is crossed: below this threshold, DNA loci are independent, above they colocalize, as a result of a thermodynamic phase transition [13, 31]. In fact, as we know from thermodynamics law, the equilibrium state of a system must correspond to a free energy minimum: so, the DNA loci form a thermodynamic stable couple only when the energy gain coming from the MF “bridges” is able to compensate the entropy loss which results from colocalization. However, in our finite size system, a narrow intermediate regime, where DNA loci couples are unstable, is found as well; the existence of such a regime is indeed typical for phase transitions in finite system [32].

Figure 3 shows the thermodynamic equilibrium value of the mean square distance  $d^2$  (Figure 3(a)) and of the fraction of colocalized DNA loci,  $P$  (Figure 3(b)), as function of the interaction energy  $E$  (here we fixed  $c = 0.2\%$  and  $n_0 = 24$ , as before). While for  $E \lesssim E^* = 1.75$  the random values for  $d^2$  and  $P$  were measured ( $d^2 \sim 40\%$  and  $P \sim 0\%$ , orange horizontal lines), as soon as  $E \gtrsim E^*$ ,  $d^2$  suddenly falls down to 0%, and, correspondingly,  $P$  increases to 100% (blue horizontal lines). In the crossover region around  $E^*$  (defined by the criterion  $P(E^*) = 50\%$ ), intermediate values for  $P$  and  $d^2$  are found, since DNA loci couples are continuously formed and disrupted.

As we explain at the beginning of this section, in the thermodynamic limit (i.e., with infinitely large system),  $E^*$  would mark the transition point between two phases (a “Brownian Phase” below it, a “Colocalization Phase” above), and a power-law behaviour would be found around  $E^*$ , for the order parameters  $d^2(E)$  and  $P(E)$  [32]. We fitted the quantitative data from the simulations, and we observed that power-law functions are also compatible with  $P(E)$  and  $d^2(E)$  for  $E \sim E^*$  in our finite-sized system (see dashed gray lines in Figures 3(a) and 3(b)):

$$d^2(E) = d_{\text{rand}}^2 - \left( \frac{E - E_0}{E_{d^2}} \right)^\alpha, \quad (5)$$

$$P(E) = \left( \frac{E - E_0}{E_P} \right)^\alpha,$$

where  $d_{\text{rand}}^2 \sim 40\%$ ,  $E_0 = 1.7 \text{ kT}$ ,  $E_{d^2} = 1.31 \text{ kT}$ ,  $E_P = 0.16 \text{ kT}$ , and  $\alpha = 0.43$ .

Since, in a cellular process, the time needed to accomplish each task is also a very important factor, we studied the behaviour of  $\tau$  (see formula (3)), which represents the characteristic time scale to reach the thermodynamic equilibrium (and colocalization, above threshold values of  $(c, E, n_0)$ ) as function of  $E$  (Figure 3(c)). An increasing behaviour is observed due to the slower DNA loci diffusion when the MF-BS bonds get stronger. The “jump” at  $E^*$  signals the phase transition occurrence. For fitting we used power-law functions (superimposed gray lines), since this is the expected behaviour for the system relaxation time near a transition point, in the thermodynamic limit. However, in our finite-sized case, we found that exponential functions can fit the data as well.

We also varied all the three parameters  $(c, E, n_0)$ , in order to find the 3D phase diagram which is plotted in Figure 4. The yellow circles mark the transition points between the “Brownian” and the “Colocalization” phase. It is shown that, as predicted by the approximate calculation of  $P$  (see (4)), the intensity of the effective attraction depends as well on MF concentration  $c$  and BS number  $n_0$ ; hence, DNA loci colocalization is also triggered when threshold values in  $c$  and  $n_0$  are passed. A power-law surface fits the data well (see [33]): we found  $cE^\gamma n_0^\delta \sim \text{const}$  with  $\gamma = 4$  and  $\delta = 1.1$ . We can conclude that colocalization is possible only if a triplet of parameters  $(c, E, n_0)$  which is above this surface is given.

### 3. Conclusions

We presented a very general mechanism which can account for different phenomena involving spatial chromatin organization. The mechanism we envisaged can produce the association between two DNA loci, without the intervention of any molecular motor able to actively move them (e.g., actin/myosin system), since the energy needed is provided by the surrounding thermal bath. The only required ingredients are those emerged in several experiments (see Section 1): say, the presence of diffusive molecules (Molecular Factors, MFs) which are able to bind specific sites (Binding Sites, BSs) within the DNA loci (Figure 1).

We carried out a detailed dissection of a schematic but quantitative model, both in its thermodynamic and dynamical aspects, by means of extensive Monte Carlo computer simulations.

We found that the DNA loci “cross-talk” and the consequent colocalization are enabled only when the MF concentration  $c$ , the BS number  $n_0$ , and the MF-BS interaction energy  $E$  are above specific threshold values, satisfying the relation  $cE^\gamma n_0^\delta = \text{cost}$  ( $\gamma = 4$  and  $\beta = 1.1$ ).

One important consequence of this result is that a cell can trigger or release pairing, for example, simply by upregulating/downregulating the MF production (i.e., by tuning  $c$ ) or by means of chromatin modification (i.e., by changing  $E$ ). Moreover, while these threshold effects can explain the results of several experiments (e.g., the importance of mediator concentration, see Section 1 and [13, 31]), they are prone to be quantitatively tested, for example, by means of BS deletions. Actually, for the first time, we introduced a model which describes such phenomena in a quantitative way: we are able to predict the likelihood of pairing and its probability distribution, once the controlling parameters ( $c, E, n_0$ ) are given.

Understanding the pairing dynamics is interesting, especially because several experiments are currently investigating it. Importantly, we measured the time scale  $\tau$  needed to reach full colocalization, as function of the system parameters. We showed that  $\tau(E)$  is an increasing function; analogously, it can be shown that  $\tau$  always increases with  $c$  and  $n_0$  as well, for the same physical reasons (see above and [13, 33]). This results in a nontrivial, counterintuitive prediction: if one (or more) of the control parameters ( $c, E, n_0$ ) is reduced (e.g.,  $n_0$  could be reduced by DNA deletions,  $E$  by chemical manipulations) at a level still above the threshold for pairing, pairing not only would occur, but it would be speeded up.

We think that the general mechanism we envisaged here, with its thermodynamic robust roots, can apply to lots of cellular processes involving DNA loci spatial organization (e.g., meiosis [33], XCI [13], chromosome territories [12], etc.) and can mark very relevant progress in their comprehension.

### 4. Materials and Methods

For MC computer simulations we used a lattice with  $L = 32$ , and so with dimensions  $L_x = 64$ ,  $L_y = 32$ , and  $L_z = 32$ , in units of  $d_0$ , the lattice spacing constant.

In this lattice, we represented DNA segments as directed polymers. Moreover, in order to reduce boundary effects, periodic boundary conditions were imposed [30]. These choices allowed faster simulations, though they do not affect the results we discussed: if such constraints are released, the free energy minimization, above specific thresholds in ( $c, E, n_0$ ), would determine the DNA loci colocalization as well.

By referring to the known case of CTCF binding in *Tsix/Xite* region on X chromosome [22, 29], we assume that the order of magnitude of  $d_0$  corresponds to the linear dimension of  $\sim 20$  bp (the length of a CTCF binding site).

The simulated DNA loci are formed by 32 beads. However, we also checked that our results remain essentially unchanged in larger lattices ( $L$  as large as 128) and with longer DNA loci (up to 128 beads).

In the implementation of MC stochastic dynamics, we considered the probability for a particle to move to an empty next neighboring site, proportional to the Arrhenius factor  $\exp(-\Delta\mathcal{H}/kT)$ ,  $\Delta\mathcal{H}$  being the energy change due to the move,  $k$  the Boltzmann constant and  $T$  the room temperature [30, 32]. To measure  $P$ , the fraction of colocalized DNA loci, we defined as “colocalized” two DNA segments at a distance less than the 10% of the lattice linear size  $L$ .

A rough calculation can be made to estimate the molecular concentration in real nuclei, corresponding to the MF volume concentrations  $c$ . Since in our model the number of molecules per unit volume is  $c/d_0^3$ , the molar concentration is  $\rho = c/(d_0^3 \mathcal{N}_A)$ , where  $\mathcal{N}_A$  is the Avogadro number. Provided that  $d_0 \sim 10$  nm (see above), we obtain that the typical nuclear protein concentration,  $\rho \sim 1 \mu\text{mol/L}$ , corresponds to  $c \sim 0.1\%$ .

Recent experiments have proved that human DNA loci, while at long times, show a constrained motion, Brownianly diffuse at short times [34]. The order of magnitude of their typical short time diffusion constant is  $D = 1 \mu\text{m}^2/\text{h}$  [34]. For sake of simplicity, we took the diffusion constant of a free polymer (i.e., with  $E = 0$ ) in our lattice equal to  $D$ ; this gives the conversion factor from MC unit time to real time (1 Monte Carlo lattice sweep  $\equiv 30$  microseconds).

### Acknowledgment

This work has been supported by the INFN-CINECA Supercomputing Project 2008.

### References

- [1] T. Misteli, “Protein dynamics: for nuclear architecture and gene expression,” *Science*, vol. 291, no. 5505, pp. 843–847, 2001.
- [2] W. de Laat and F. Grosveld, “Spatial organization of gene expression: the active chromatin hub,” *Chromosome Research*, vol. 11, no. 5, pp. 447–459, 2003.
- [3] K. J. Meaburn and T. Misteli, “Cell biology: chromosome territories,” *Nature*, vol. 445, no. 7126, pp. 379–781, 2007.
- [4] P. Fraser and W. Bickmore, “Nuclear organization of the genome and the potential for gene regulation,” *Nature*, vol. 447, no. 7143, pp. 413–417, 2007.

- [5] J. Dekker, "Gene regulation in the third dimension," *Science*, vol. 319, no. 5871, pp. 1793–1794, 2008.
- [6] T. Misteli, "Cell biology: nuclear order out of chaos," *Nature*, vol. 456, no. 7220, pp. 333–334, 2008.
- [7] T. Takizawa, K. J. Meaburn, and T. Misteli, "The meaning of gene positioning," *Cell*, vol. 135, no. 1, pp. 9–13, 2008.
- [8] R. I. Kumaran, R. Thakar, and D. L. Spector, "Chromatin dynamics and gene positioning," *Cell*, vol. 132, no. 6, pp. 929–934, 2008.
- [9] D. Zickler and N. Kleckner, "The leptotene-zygotene transition of meiosis," *Annual Review of Genetics*, vol. 32, pp. 619–697, 1998.
- [10] J. L. Gerton and R. S. Hawley, "Homologous chromosome interactions in meiosis: diversity amidst conservation," *Nature Reviews Genetics*, vol. 6, no. 6, pp. 477–487, 2005.
- [11] D. Zickler, "From early homologue recognition to synaptonemal complex formation," *Chromosoma*, vol. 115, no. 3, pp. 158–174, 2006.
- [12] M. Nicodemi and A. Prisco, "Thermodynamic pathways to genome spatial organization in the cell nucleus," *Biophysical Journal*, vol. 96, no. 6, pp. 2168–2177, 2009.
- [13] A. Scialdone and M. Nicodemi, "Mechanics and dynamics of X-chromosome pairing at X inactivation," *PLoS Computational Biology*, vol. 4, no. 12, article e1000244, 2008.
- [14] S. Cai, C. C. Lee, and T. Kohwi-Shigematsu, "SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes," *Nature Genetics*, vol. 38, no. 11, pp. 1278–1288, 2006.
- [15] R. Schneider and R. Grosschedl, "Dynamics and interplay of nuclear architecture, genome organization, and gene expression," *Genes & Development*, vol. 21, no. 23, pp. 3027–3043, 2007.
- [16] H. Jing, C. R. Vakoc, L. Ying, et al., "Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus," *Molecular Cell*, vol. 29, no. 2, pp. 232–242, 2008.
- [17] A. Wutz and J. Gribnau, "X inactivation Xplained," *Current Opinion in Genetics and Development*, vol. 17, no. 5, pp. 387–393, 2007.
- [18] P. Avner and E. Heard, "X-chromosome inactivation: counting, choice and initiation," *Nature Reviews Genetics*, vol. 2, no. 1, pp. 59–67, 2001.
- [19] M. Nicodemi and A. Prisco, "Symmetry-breaking model for X-chromosome inactivation," *Physical Review Letters*, vol. 98, no. 10, Article ID 108104, 2007.
- [20] M. Nicodemi and A. Prisco, "Self-assembly and DNA binding of the blocking factor in X chromosome inactivation," *PLoS Computational Biology*, vol. 3, no. 11, article e210, 2007.
- [21] N. Xu, C.-L. Tsai, and J. T. Lee, "Transient homologous chromosome pairing marks the onset of X inactivation," *Science*, vol. 311, no. 5764, pp. 1149–1152, 2006.
- [22] N. Xu, M. E. Donohoe, S. S. Silva, and J. T. Lee, "Evidence that homologous X-chromosome pairing requires transcription and Ctf protein," *Nature Genetics*, vol. 39, no. 11, pp. 1390–1396, 2007.
- [23] L.-F. Zhang, K. D. Huynh, and J. T. Lee, "Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing," *Cell*, vol. 129, no. 4, pp. 693–706, 2007.
- [24] M. Doi and S. Edwards, *The Theory of Polymer Dynamics*, Clarendon Press, Oxford, NY, USA, 1986.
- [25] S. J. Maerkl and S. R. Quake, "A systems approach to measuring the binding energy landscapes of transcription factors," *Science*, vol. 315, no. 5809, pp. 233–237, 2007.
- [26] A. V. Morozov, J. J. Havranek, D. Baker, and E. D. Siggia, "Protein-DNA binding specificity predictions with structural models," *Nucleic Acids Research*, vol. 33, no. 18, pp. 5781–5798, 2005.
- [27] U. Gerland, J. D. Moroz, and T. Hwa, "Physical constraints and functional characteristics of transcription factor-DNA interaction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 19, pp. 12015–12020, 2002.
- [28] M. Lässig, "From biophysics to evolutionary genetics: statistical aspects of gene regulation," *BMC Bioinformatics*, vol. 8, supplement 6, article S7, 2007.
- [29] M. E. Donohoe, L.-F. Zhang, N. Xu, Y. Shi, and J. T. Lee, "Identification of a Ctf cofactor, Yy1, for the X chromosome binary switch," *Molecular Cell*, vol. 25, no. 1, pp. 43–56, 2007.
- [30] K. Binder, "Applications of Monte Carlo methods to statistical physics," *Reports on Progress in Physics*, vol. 60, no. 5, pp. 487–559, 1997.
- [31] M. Nicodemi, B. Panning, and A. Prisco, "A thermodynamic switch for chromosome colocalization," *Genetics*, vol. 179, no. 1, pp. 717–721, 2008.
- [32] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Clarendon Press, Oxford, NY, USA, 1971.
- [33] M. Nicodemi, B. Panning, and A. Prisco, "The colocalization transition of homologous chromosomes at meiosis," *Physical Review E*, vol. 77, no. 6, Article ID 061913, 2008.
- [34] J. R. Chubb, S. Boyle, P. Perry, and W. A. Bickmore, "Chromatin motion is constrained by association with nuclear compartments in human cells," *Current Biology*, vol. 12, no. 6, pp. 439–445, 2002.