The colocalization transition of homologous chromosomes at meiosis

Mario Nicodemi,[1,2] Barbara Panning,[3] and Antonella Prisco[4]
[1]*Department of Physics and Complexity Science, University of Warwick, Coventry CV4 7AL, United Kingdom*
[2]*INFN, Napoli, Italy*
[3]*Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California 94143, USA*
[4]*CNR Inst. Genet. and Biophys. "Buzzati Traverso," Via P. Castellino 111, Napoli, Italy*
(Received 7 November 2007; revised manuscript received 28 March 2008; published 17 June 2008)

Meiosis is the specialized cell division required in sexual reproduction. During its early stages, in the mother cell nucleus, homologous chromosomes recognize each other and colocalize in a crucial step that remains one of the most mysterious of meiosis. Starting from recent discoveries on the system molecular components and interactions, we discuss a statistical mechanics model of chromosome early pairing. Binding molecules mediate long-distance interaction of special DNA recognition sequences and, if their concentration exceeds a critical threshold, they induce a spontaneous colocalization transition of chromosomes, otherwise independently diffusing.

DOI: 10.1103/PhysRevE.77.061913 PACS number(s): 87.14.G−, 64.60.Cn, 82.39.Rt, 87.15.A−

At the onset of meiosis—the specialized cell division producing cells (gametes) with half the genetic content of the mother—homologous chromosomes in the cell nucleus associate in couples via a process involving a long-distance recognition of "self" followed by a gradual alignment. Such a process is crucial for preventing fertility problems, birth defects, and cancer, but it remains one of the most mysterious aspects of meiosis [1–4]. Although substantial variation among organisms exists, some of the essential elements involved in pairing are conserved. Yet we still do not know how they generate self-recognition and colocalization. Starting from recent experimental hints on the molecular basis, we discuss here a statistical mechanics model explaining the mechanical features of early pairing and their active control by the cell.

At meiosis, paired homologs systematically exchange DNA strands in a process known as genetic recombination, which is initiated by DNA double-strand breaks (DSBs). A long-standing hypothesis proposes, thus, that pairing is DSB-related, for instance by a direct physical interactions between DNA homologous duplexes. While alternative scenarios were discussed (see, e.g., [5]), recent experiments have changed such a picture: DSBs are required for attaining full alignment, but the early stages of homolog pairing appear to be DSBs independent in many organisms (as seen in fungi, plants, and higher animals [1–4,6]). In particular, recent results on model organisms such as *C. elegans* and *Drosophila* shed new light on DSB-independent pairing mechanisms [7–10]. In *C. elegans*, homologs colocalization is known to depend on special recognition regions at their ends (telomeric regions), known as "pairing centers" (PCs) [11–13]. At the early stages of meiosis, in early prophase I, special DNA binding proteins, HIM/ZIM (in the class of "zinc fingers"), localize to their corresponding PC and mediate their pairing [14]. Similarly, in *Drosophila,* clusters of a 240-base-pair repeat sequence at a specific location on the X and Y chromosomes, and analogous sites on nonsex chromosomes, act as a pairing center (see Refs. [2,3]); and special proteins, such as MNM and SNM (also a zinc finger), binding X-Y and nonsex chromosome pairing sites at prophase I, are necessary to observe pairing [15]. Another "universal" feature of prophase I is telomere tethering to the nuclear envelope, which is considered to be intimately related to pairing [1–4]. Such a picture, where early recognition is mediated by special chromosomal regions ("pairing centers") in interaction with a set of specific proteins, is interestingly found across organisms [1–4]. Nevertheless, the likelihood of random contacts of homologous recognition elements is negligible, so the "Maxwell's demon" [11] responsible for homology sensing and colocalization remains elusive.

Here we discuss a statistical mechanics model that incorporates the minimal physical elements revealed by experiments, i.e., DNA pairing sequences and molecules binding them. These elements are shown to be sufficient to chromosome colocalization as binding molecules can induce an effective attraction between DNA recognition sequences: when the concentration of molecules, or DNA chemical affinity, exceeds a given threshold, a phase transition occurs (in a finite-sized system) and homologs are spontaneously joined together, otherwise they move independently. Our model is very schematic and many complexities are not considered, yet such a "thermodynamic switch" is robust, irrespective of its ultimate molecular and biochemical basis; its aim is to delineate a simple conceptual framework and the key ingredients necessary to the early events in colocalization.

Model. Our model (see right panels in Fig. 2) consists of a pair of identical chromosome segments, involved in recognition, and a concentration, c, of molecular factors having a chemical affinity, E_X, for them. For the sake of simplicity, we ignore the rest of the chromosomes, and DNA segments are described as directed polymers in resemblance to DNA telomeres tethering to the nuclear envelope at meiosis [16]. In our Monte Carlo computer simulations [17], molecules and polymers diffuse in a cubic lattice with spacing d_0 (of the order of the unknown molecular factors length) and linear sizes $L_x=2L$, $L_y=L$, and $L_z=L$ (in units of d_0). DNA segments are dealt with as a string of L nonoverlapping "beads" diffusing under the constraint that two proximal beads must be on next or next-nearest-neighboring sites on the lattice (on each vertex, no more than one particle can be present at a given time). Chromosomal beads interact, via an effective energy E_X, with neighboring molecules which, in turn, can

 ©2008 The American Physical Society

bind both polymers at the same time, as suggested by a number of mediating proteins discovered to date, e.g., ZIM/HIM in *C. elegans* having many DNA binding domains. In order to stress that "weak" biochemical bonds are sufficient to hold in place whole chromosomes, here we mostly focus on the example in which E_X is of the order of a weak hydrogen bond, say 3 kJ/mol or $E_X=1.2kT$ at room temperature [18].

Finally, we shall also briefly discuss the case in which molecular mediators can form complexes, i.e., interact with nearest-neighbor molecules via a reciprocal affinity E_0, but now we set $E_0=0$. In our simulations, the "beads" of the chromosomal segments initially form two straight vertical lines, at a distance L from each other, and molecular mediators are randomly distributed. The size of our lattice is $L=16$ (we checked our results for L as large as 128) with periodic boundary conditions. Averages are over up to 2048 runs from different initial configurations, and time is given in units of Monte Carlo lattice sweeps [17]. The probability of a particle moving 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 is the Boltzmann constant, and T is the temperature [18,19]. For the reaction collision rate, r_0 (depending on the nature of the molecular factors and of the surrounding viscous fluid), we use $r_0=30\ \mathrm{s}^{-1}$, a typical value in biochemical kinetics.

Colocalization dynamics. Although molecular mediators can have a weak affinity for chromosomes, say $E_X=1.2kT$, they can collectively induce an effective attraction between them: if c is above a threshold value, the effective interaction is strong enough to result in chromosome colocalization. This is shown by the average square distance between the two chromosomal strings (relative to the system linear size L): $d^2(t)=\Sigma_{z=1}^{N}\langle r^2(z,t)\rangle/L^2N$, where N is the number of beads in each string (here $N=L$) and $\langle r^2(z,t)\rangle$ is the average, over different realizations, of the square distance of the beads at "height" z at time t on the two chromosomes. Figure 1 shows $d^2(t)$ [which starts from $d^2(0)=100\%$] as a function of time, t, for two relevant values of c. For instance, for $c=2.5\%$, $d^2(t)$ at long times converges almost to zero: this is the sign that the two "chromosomes" have colocalized; a typical picture of the system final state is shown in the lower right panel in Fig. 2 (real DNA recognition sequences could be locally bound without such a precise alignment as in our simple model). Conversely, when particle density is small enough (say for $c=0.3\%$), $d(t)$ saturates to a finite plateau value of the order of the system size (around 40% of L^2), corresponding to the average distance of two independent, nonoverlapping strings undergoing Brownian motion in a box of size L, a typical configuration being shown in Fig. 2, upper right panel.

As seen in Fig. 1, after an initial Brownian linear behavior in t, $d^2(t)$ approaches its equilibrium value approximately exponentially in t, $d^2(t)\propto\exp(-t/\tau)$ (superimposed fits in Fig. 1). From this relation, we extract the equilibration time scale, τ. As shown in the inset of Fig. 1, τ is approximately a power law in c (superimposed fits in Fig. 1, inset): $\tau\sim c^{\alpha}$, with an exponent α close to 1 (smoothly depending on E_X and E_0). τ is also a function of E_X and E_0. When E_X rises, molecules tend to cluster more tightly around each polymer segment making their dynamics slower and decreasing the chances of a random locking encounter between them. Conversely, by increasing E_0, molecules can aggregate facilitating segment coupling via cooperative-like behaviors. So, τ increases with E_X and decreases with E_0 (see Fig. 1, inset).

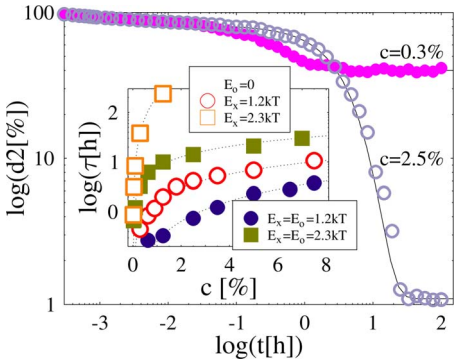

FIG. 1. (Color online) Main panel shows the evolution of the square distance between "chromosome recognition sequences," $d^2(t)$ (expressed as a fraction of the system size, L), as a function of time, t (in $\log_{10}$ scale), for the given values of the concentration of molecular mediators, c (here the molecule-chromosome and molecule-molecule affinities are, respectively, $E_X=1.2kT$ and $E_0=0$). Chromosomes are initially at a distance L, i.e., $d^2(0)=100\%$. At long times $d^2(t)$ plateaus to its equilibrium value approximately exponentially (superimposed fit), which for $c=0.3\%$ corresponds to the expected average square distance between independent randomly diffusing directed strings, ~40%; for $c=2.5\%$, instead, $d^2(t)$ collapse to almost zero, signaling that the two chromosomes have colocalized. Inset shows (in a range of values of E_X and E_0) that the characteristic time to approach equilibrium, τ (derived from exponential fits), increases as a function of c. The superimposed fits are power laws.

Phase diagram. Colocalization is attained when c is higher than a given threshold, as shown by the equilibrium value, $d^2(c)$, of their distance plotted as a function of c in Fig. 2, left panel. Above a critical point c_{tr} [approximately defined by the inflection point of $d^2(c)$, e.g., $c_{\mathrm{tr}}\simeq 0.7\%$ for $E_X=1.2kT$], $d^2(c)$ rapidly goes to zero and the system is found in the regime where chromosomes are tightly colocalized, the "colocalization phase." Conversely, when c is small enough, say below c_{tr}, $d^2(c)$ has the same value found for two noninteracting Brownian strings; the effective attraction is too small and chromosomes float away one from the other. This is the "random phase" where chromosomes are independent. The phase transition, in this finite-size system, occurs when entropy loss due to string colocalization is compensated by particle energy gain as they bind both strings. Actually, the transition is found in a broad region of the (E_X,c) plane, as shown in Fig. 3, where the system phase diagram is plotted in a range of typical biochemical values of binding energies E_X.

The phase diagram shows that any binding energy above a threshold of the order of a weak h-bond would work to attain chromosome pairing; the higher E_X, the smaller the required mediator concentration, c. So, molecules binding specific DNA sequences (e.g., 10–20 bases long) with comparatively higher overall energies (say $20kT$) would fit well in our picture of colocalization. Specificity of colocalization among many a chromosome pair could be, indeed, obtained by sets

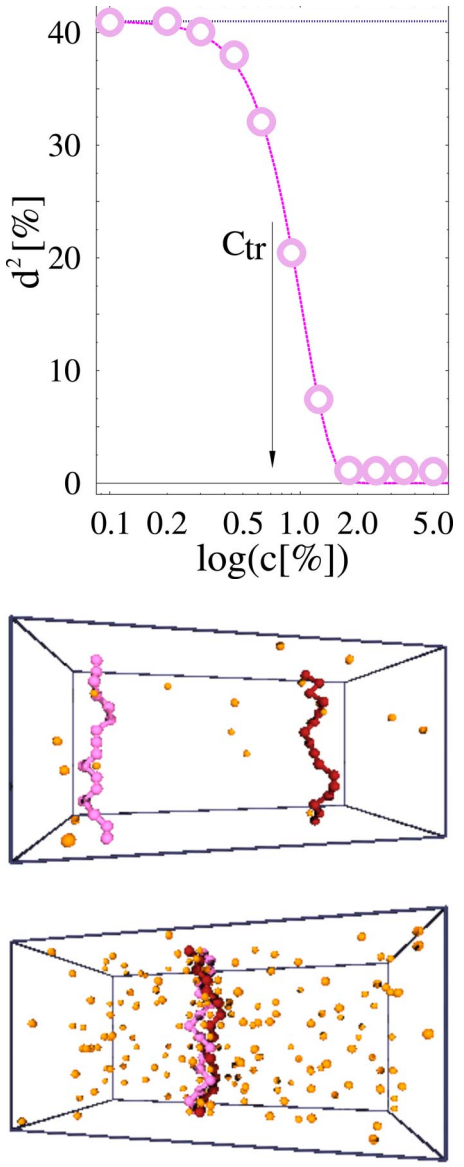

FIG. 2. (Color online) Left panel The equilibrium average chromosome square distance, d^2, from Fig. 1 ($E_X=1.2kT$, $E_0=0$) is shown as a function of c (in $\log_{10}$ scale): for $c<c_{tr}\simeq 0.7\%$, d^2 approaches values as big as the system size and chromosomes are randomly and independently diffusing (horizontal dotted line gives the independent diffusion value $d^2\sim 40\%L^2$); for $c>c_{tr}$, d^2 rapidly decays to zero, showing that they have colocalized. Around c_{tr}, there is a narrow crossover regime where chromosomes are only transiently colocalizing. Right panels show pictures of typical equilibrium configurations, in the "independent" (upper panel, $c=0.3\%$) and "colocalized' phase" (lower panel, $c=2.5\%$). Molecular mediators are pictured as orange floating particles.

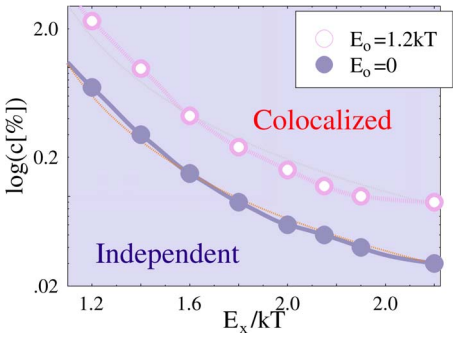

FIG. 3. (Color online) The system phase diagram is plotted in the (E_X,c) plane (the c axis is in $\log_{10}$ scale). The transition line, $c_{tr}(E_X)$, delimiting the phase where chromosomes move independently from the phase where they colocalize, is marked by filled circles in the case $E_0=0$. For comparison, the transition line is also plotted in the case $E_0=1.2kT$ (empty circles). Fits superimposed to the two transition lines are inverse power laws (see text).

of molecules binding only specific homologs. From experiments in *C. elegans* we know, for instance, that HIM-8 proteins binds only X chromosomes, while ZIM-2 only V chromosomes. The phase diagram also suggests a different path to attain colocalization: for a given physiological concentration of binding molecules, c; the cell could just chemically act on the DNA recognition sequences to increase their affinity, E_X, to molecules: this would induce spontaneous colocalization as well (see Fig. 3). In the presence of a nonzero reciprocal affinity, E_0, when molecular mediators can form aggregates [24,25], the phase diagram topology is not altered, but the location of the transition line $c_{tr}(E_X)$ moves toward higher concentrations (see Fig. 3).

The transition from the colocalized to the independent state may be understood by a simplified calculation. Suppose the two polymers are in the colocalization regime and denote their trajectory through the system $\mathbf{r}_1(z)$ and $\mathbf{r}_2(z)$. Their energy E can be approximated as the sum of the bending energy and the effective interaction potential, V. For the sake of simplicity, we neglect self-repulsion and expand V to second order around its minimum. So we have [20]

$$E[\mathbf{r}_1,\mathbf{r}_2]=\int_0^L dz\left\{\frac{b}{2}[(\mathbf{r}_1')^2+(\mathbf{r}_1')^2]+\frac{v}{2}(\mathbf{r}_1-\mathbf{r}_2)^2\right\}, \quad (1)$$

where $\mathbf{r}_1'=d\mathbf{r}_1(z)/dz$, b is the tilt modulus, and $v=V''(0)$. Assume that a polymer requires a distance $z=l$ to span its typical transverse fluctuation, Δr. From Eq. (1), the energy of a segment l is $E(l)\simeq(bl^{-1}+vl)\Delta r^2/2$. The equilibrium value of l, l_0, is obtained by minimization: $\partial E/\partial l|_{l_0}=0$. This gives $l_0=\sqrt{b/v}$, corresponding to an energy $E(l_0)=\sqrt{bv}\Delta r^2$. At equilibrium, the equipartition theorem implies that $E(l_0)=kT$, so we get $\Delta r^2=kT/\sqrt{bv}$. We can apply a criterion in the manner of Lindemann to establish the maximal value of Δr^2 above which colocalization is no longer sustainable against thermal fluctuations,

$$\Delta r^2\simeq f_0^2 d^2, \quad (2)$$

where d is a scale of the order of the range of $V(r)$ and f_0 is the Lindemann constant, say $f_0\sim O(10^{-1})$. Equation (2) then implies $kT/\sqrt{bv}\sim f_0^2 d^2$. As the interaction potential is generated by molecules (taken from a concentration c) binding either polymer (the binding energy being $2E_X$), at a first rough approximation we have $v\propto cE_X$, ensuring the physical constraint whereby v vanishes when either $c\to 0$ or $E_X\to 0$. More generally, the relation between v, c, and E_x can involve more complex powers (e.g., $v\propto c^\alpha E_x^\beta$). By substitution in Eq.

(2), we get a power-law expression, with an exponent ν (e.g., $\nu=\beta/\alpha$), for the colocalization and independence transition line in terms of c and E_X,

$$c_{\mathrm{tr}}(E_X) \sim E_X^{-\nu}. \tag{3}$$

Figure 3 shows an approximate power-law fit, with $\nu=2$, to the numerically derived transition line $c_{\mathrm{tr}}(E_X)$ in both cases $E_0=0$ and $E_0=1.2kT$. The fit can be further improved by introducing a minimal threshold energy $E^* \sim 0.8kT$, below which no transition is possible: $c_{\mathrm{tr}}(E_X) \sim (E_X - E^*)^{-\nu}$.

Conclusions. In summary, we have described a general colocalization mechanism whereby specific regions of a pair of chromosomes can recognize each other and align. Under this model, physical juxtaposition of recognition sequences is mediated by sequence-specific molecular factors binding DNA via weak, nonpermanent, biochemical interactions. We showed that by tuning the concentration of molecular mediators or DNA affinity to them, the cell controls a switch for pair formation and release, a mechanism having general and robust roots [19] in a thermodynamics phase transition occurring in the system. In an alternative scenario, proteins able to bind only one DNA binding site, but linked by protein-protein interactions, could also act for colocalization; as a pair of linked proteins can be represented, in the model we discuss here, via a single molecular mediator; we expect that the thermodynamics picture is unchanged.

While other biological processes can help pairing, our mechanisms explain how the minimal "ingredients" discovered in experiments (i.e., soluble molecules and DNA binding sites) can drive early pairing and are sufficient for it. As our model provides the required general thermodynamic grounds, it is very schematic [21]: for instance, only two chromosomal segments are considered. Simulations with many pairs of "chromosomes" show longer time scales to approach equilibrium, as expected in a crowded environment. Yet the universality of the statistical mechanics phenomenon is not affected and the overall thermodynamic phase diagram is unchanged. As discussed in, e.g., [5], one of the advantages of a pairing initially based on nonpermanent interaction, as in the scenario discussed here, is to prevent ectopic association between nonhomologous and avoid topologically unacceptable entanglement of chromosomes, leaving space for adjustments. Early recognition sequences can later seed a correct chromosome-wide homologous pairing. Other chromosome-unspecific molecular mediators could help the overall process.

Prone to experimental tests (e.g., threshold in c, dynamics, ...), our model can be used to predict the effects of DNA chemical modifications (changes to E_X) and deletions (insertions) [decrease (increase) the number, L, of DNA binding sites). Many other cell processes involve the organization of chromosomes in nuclear space (e.g., X pairing at the onset of X-inactivation [22–25], chromosome architecture [26]), and the mechanism described here could be relevant to those cases as well.

[1] D. Zickler and N. Kleckner, Annu. Rev. Genet. **32**, 619 (1998).
[2] J. L. Gerton and R. Scott Hawley, Nat. Rev. Genet. **6**, 477 (2005).
[3] D. Zickler, Chromosoma **115**, 158 (2006).
[4] G. S. Roeder, Genes Dev. **11**, 2600 (1997).
[5] N. Kleckner and B. M. Weiner, Cold Spring Harb. Symp. Quant. Biol. **LVIII**, 553 (1993).
[6] S. Tessé *et al.*, Proc. Natl. Acad. Sci. U.S.A. **100**, 12865 (2003).
[7] A. F. Dernburg *et al.*, Cell **94**, 387 (1998).
[8] K. S. McKim *et al.*, Science **279**, 876 (1998).
[9] A. J. MacQueen *et al.*, Genes Dev. **16**, 2428 (2002).
[10] J. Vazquez *et al.*, Curr. Biol. **12**, 1473 (2002).
[11] A. J. MacQueen *et al.*, Cell **123**, 1037 (2005).
[12] K. S. McKim *et al.*, Genetics **120**, 987 (1988).
[13] A. M. Villeneuve, Genetics **136**, 887 (1994).
[14] C. M. Phillips and A. F. Dernburg, Dev. Cell **11**, 817 (2006).
[15] S. E. Thomas *et al.*, Cell **123**, 555 (2005).
[16] Telomeres tethering to the envelope and weak interaction energies reduce topological entanglement. Simulations of two "longer" chromosomes including binding and nonbinding segments give results analogous to those reported here. A model including many pairs of envelope-tethered chromosomes has a similar phase diagram.
[17] K. Binder, Rep. Prog. Phys. **60**, 487 (1997).
[18] J. D. Watson *et al.*, *Molecular Biology of the Gene* (Benjamin Cummings, New York, 2003).
[19] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Clarendon, Oxford, 1971).
[20] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, Oxford, 1987).
[21] M. Nicodemi, B. Panning, and A. Prisco, Genetics **179**, 717 (2008).
[22] N. Xu, C.-L. Tsai, and J. T. Lee, Science **311**, 1149 (2006).
[23] C. P. Bacher *et al.*, Nat. Cell Biol. **8**, 293 (2006).
[24] M. Nicodemi and A. Prisco, Phys. Rev. Lett. **98**, 108104 (2007).
[25] M. Nicodemi and A. Prisco, PLOS Comput. Biol. **3**, e210 (2007).
[26] C. Lanctôt *et al.*, Nat. Rev. Genet. **8**, 104 (2007).