Lattice animal model of chromosome organization

Balaji V. S. Iyer and Gaurav Arya*
Department of NanoEngineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0448, USA
(Received 23 June 2011; published 12 July 2012)

Polymer models tied together by constraints of looping and confinement have been used to explain many of the observed organizational characteristics of interphase chromosomes. Here we introduce a simple lattice animal representation of interphase chromosomes that combines the features of looping and confinement constraints into a single framework. We show through Monte Carlo simulations that this model qualitatively captures both the leveling off in the spatial distance between genomic markers observed in fluorescent *in situ* hybridization experiments and the inverse decay in the looping probability as a function of genomic separation observed in chromosome conformation capture experiments. The model also suggests that the collapsed state of chromosomes and their segregation into territories with distinct looping activities might be a natural consequence of confinement.

DOI: 10.1103/PhysRevE.86.011911 PACS number(s): 87.16.Sr, 87.16.af, 87.16.aj

I. INTRODUCTION

Eukaryotic genomes are organized in a hierarchical manner in the cell nucleus: The DNA is first associated with histone proteins to form the chromatin fiber, which then undergoes further "rounds" of folding to yield chromosomes [1]. The higher-order folding of chromatin into chromosomes plays many important biological functions beyond its traditional role in DNA packaging. Specifically, the formation of chromatin loops dictates interactions between distant portions of the genome to orchestrate and regulate DNA transcription [2] and recombination [3]. Further, the spatial location of genes within chromosome territories and the folding of proximal chromatin has important consequences on gene activity and function [4,5]. However, due to limitations in visualizing chromatin fibers *in vivo* [6], a comprehensive knowledge of the higher-order organization of chromosomes and the physical principles governing it remains elusive.

Increasingly, innovative experiments in conjunction with polymer models are being employed to indirectly probe the organization of interphase chromosomes. The fluorescence *in situ* hybridization (FISH) experiments measure the Euclidean distance between genomic markers as a function of their separation (in bp) [7]. These measurements reveal an initial power-law rise in the spatial distance with separation followed by a "levelling off" beyond ∼1–10 Mbp, where the distance becomes independent of separation [3,8]. The earliest models considered chromosomes as confined Gaussian polymer chains [9,10] to explain the rise and levelling off in spatial distances. However, an alternative explanation for the levelling off based on incorporation of loops was found to be more consistent with the observation of looped structures in light microscopy experiments [11]. This proposition led to the formulation of the random walk giant loop (RWGL) model [7] and increasingly refined models like the multiloop subcompartment (MLS) [12] and random loop (RL) models [13]. The chromosome conformation capture (3C) experiments [14,15], on the other hand, measure the frequency with which distant regions of the genome interact with each other to form loops. Recent measurements on human chromosomes [15] indicate that the probability of loop formation $P(s)$ decays roughly inversely with the loop size s in the ∼0.5- to 7-Mbp range, suggesting that human chromosomes might be organized as a fractal network of loops within this range. A fractal globule (FG) model was proposed to explain this decay [15]. In this model, the chromosome was treated as a linear polymer chain that is suddenly compressed and prevented from equilibrating. This fractal nature is also supported by small-angle neutron scattering [16] and tracer diffusion [17] experiments.

While the loop and fractal themes independently support observations from FISH and 3C, respectively, neither theme can simultaneously explain results from both experiments. For instance, the loop models do not capture the observed decay in loop probability due to their assumption of fixed loops in the case of RWGL and MLS models or equal *a priori* probabilities of loop formation in the case of RL model while the fractal-based models do not capture the levelling off in spatial distances at distances smaller than the confinement length scale. Recently, two attempts have been made to reconcile observations from FISH and 3C measurements within a single framework. In one model, the looping probabilities in the RL model were modified by invoking diffusion controlled colocalization of loops [18], though the decay scaling in the looping probability appears to be very sensitive to the imposed lifetime of loops. In another model, curvature defects or kinks were introduced into the chromatin fiber treated as a semiflexible chain [19] to yield the 3C looping probabilities. This model also captured the power-law rise in spatial distances, though their eventual levelling off was not captured. Furthermore, the loop and fractal themes invoke different underlying mechanisms to capture looping and segregation. For example, the loop models [8,18] enforce looping through the use of transient contact potentials between random locations on the polymer chain and the topological constraints between loops are expected to drive segregation [20,21]. In contrast, strong confinement is required in the fractal models to force contacts (loops) between distant segments of a polymer chain and the kinetic entrapment of loops is supposed to give rise to segregation [15].

In this article, we introduce a lattice representation of interphase chromosomes that naturally builds into it the presence of random loops at all length scales and their fractallike organization. The model allows us to simultaneously capture

*garya@ucsd.edu

the levelling off in spatial distances and the inverse decay in looping probabilities observed in FISH and 3C measurements, respectively. We find that excluded volume interactions, which prevent loops and looping points from interpenetrating, and confinement effects, which limit the formation of highly elongated and hierarchical loops, seem to be key factors responsible for the observed features of chromosomes. The model also illustrates how simple physical constraints could give rise to complex structural effects within looped structures, such as their collapse and segregation into territories with distinct looping activities.

II. LATTICE ANIMAL MODEL

Interphase chromosomes can be modeled as linear, semiflexible polymer chains, representing the underlying chromatin fiber, tied together by constraints from looping and confinement. However, the physical properties of the chromatin fiber and the molecular mechanisms confining the chromosomes to within territories remain poorly understood [22]. Under such conditions, it is perhaps more useful to have a coarser representation of chromosomes that allows imposition of physical constraints, without the need of explicitly simulating the molecular processes leading to the constraints, for predicting qualitative scaling relations. In this article we take such an approach, whereby chromosomes are construed as 3D networks of looping points. We demonstrate that such an abstract representation can be used to understand the role of basic constraints on chromosome organization.

To model chromosomes, we adapt the lattice animal (LA) representation (Fig. 1) used previously to model percolation networks and branched polymers [23]. The chromosome is represented by a set of nodes connected to each other via links. The internal nodes, connected to >1 links, represent looping points, where distant segments of chromatin come together, whereas the end nodes, connected to one link, represent loop termini. Each link represents two chromatin fibers running in opposite directions. To keep the model simple and computationally tractable, we enforce the nodes to occupy positions on a cubic lattice with coordination number $z = 6$ and lattice spacing $b = 1$. Thus, a chromosome described by N nodes contains $N - 1$ links and has a total contour length of $L = 2 \times (N - 1)$, the factor of 2 arising from the double-stranded nature of each link. As illustrated in Fig. 1, the LA model naturally accounts for the existence of loops of different sizes, where internal nodes carry loops of sizes $s \geqslant 2$, quantized in multiples of 2. The model also importantly allows for dynamic relocation of loops through rearrangement of nodes. Such relocation of loops is equivalent to the continuous reorganization of the genome triggered by transcription and regulatory mechanisms and also mimics the ensemble averaging over multiple cells and time points conducted in the experiments.

The LA representation of chromosomes offers several other advantages. First, the existence of random loops of all length scales and their fractal organization, which are central to the RL and FG models, is naturally built into the LA representation. Second, loop sizes or loop size distribution need not be specified *a priori* but evolve as a function of the applied physical constraints. Third, the representation allows for the use of existing ideas in the rich fields of percolation networks, branched polymers, and critical phenomena [23] to bear on the problem of chromosome organization. For instance, the LA structures studied here belong to the same universality class as randomly branched polymers with annealed branch points [23], which allows for comparison of scaling exponents and use of existing free energy arguments to build appropriate physical constraints.

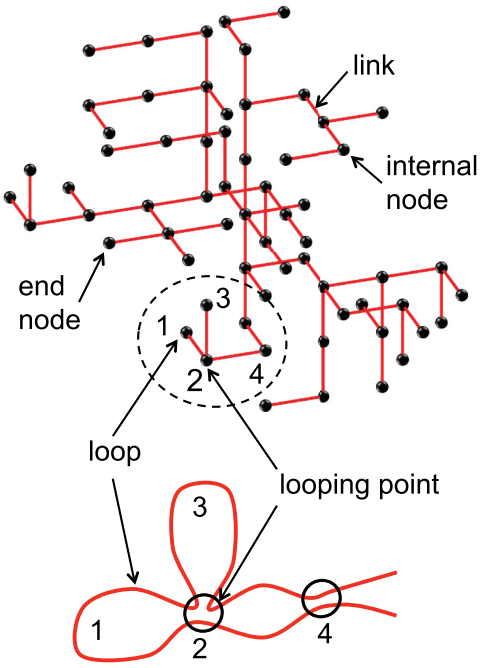

FIG. 1. (Color online) Lattice animal representation of chromosomes. The nodes, depicted via black dots, represent looping points or loop termini. The links, depicted via red (gray) lines connecting the nodes represent chromatin fibers running in opposite directions. The circled region illustrates a looping point numbered 2 carrying two loops, each of size $s = 2$, and a looping point numbered 4 carrying one loop of size $s = 6$.

So far, we have not considered any restrictions on the LA nodes. It is, however, likely that chromosome loops and looping points experience constraints from the excluded volume of the chromatin fiber and the compartmentalization of chromosomes into territories in the nucleus. There also likely exist physical and biological bounds on the size and hierarchy of loops present within chromosomes. In this study, we examine the effect of various physical constraints on the structural properties of LAs. Our aim is to determine the conditions under which the LA model reproduces the levelling off of the intergenomic spatial distances observed in FISH measurements and the specific loop size distributions observed in 3C experiments. We examine *four* different types of LAs here, which are described below.

Ideal LA. This structure represents the most basic of LAs with no imposed constraints on the nodes. Since excluded volume interactions are absent, two or more nodes are allowed to occupy the same lattice point without any energetic penalty. The total energy $U = 0$ for ideal LAs.

Self-avoiding LA. A hard sphere repulsion U_{exc} is introduced between all nodes such that when two or more nodes

occupy the same lattice point $U_{\rm exc} \to \infty$, otherwise $U_{\rm exc} = 0$ when no such overlaps exist. Physically, this excluded volume constraint implies that the chromatin loops (links) cannot interpenetrate each other, which seems reasonable given that chromosomes are averse to forming knots [15,24]. This constraint also implies that two different looping points cannot occupy the same volume, which also seems physically reasonable. The total energy of such LAs is given by $U = U_{\rm exc}$.

Spatially confined LA. Chromosomes are more or less globular in appearance and are generally localized within territories inside the nucleus [25]. One possible source for this behavior could be the entropic repulsion between chromosomes due to their loopy conformations [21]. To mimic such a spatial confinement in the LA model, we impose an energetic penalty $U_{\rm conf}$ on the radius of gyration R_g of the LAs such that highly swollen or elongated conformations are disfavored. We also enforce an excluded volume constraint on the nodes, as the confinement constraint becomes irrelevant without the former. Before implementing the confinement constraint, we obtain the radius of gyration $R_g^{'}$ of an unconstrained, self-avoiding LA. To implement confinement effects such that the LAs exhibit $R_g < R_g'$, we construct an energy function $U_{\rm conf} = k_B T \exp[\sigma_r(R_g - \alpha R_g^{'})]$, based on a cage potential used earlier for confining linear polymers [15]. In this function, R_g is the radius of gyration of the existing LA structure, $\sigma_r \equiv R_g'/z$ weights the degree of confinement according to the LA size and the lattice coordination number, and $k_B T$ (set to 1) is the thermal energy. The parameter $\alpha < 1$, the only adjustable parameter in this function, governs the strength of confinement. The smaller the α, the stronger the confinement. The rapidly rising nature of the exponentially varying confinement potential thus suppresses the formation of large and highly elongated structures, as illustrated in Fig. 2. The total energy of spatially confined LA structures is given by $U = U_{\rm conf} + U_{\rm exc}$.

Generation-number confined LA. Chromosomes are believed to be organized as a nested hierarchy of loops within loops. It is conceivable that this hierarchy does not continue indefinitely but is constrained at some level due to physical and biological limits. Such limits could arise from difficulties in bending increasingly higher-order structures of chromatin into loops, similar to the limit on the number of times one can fold a piece of paper, or from difficulties in regulating extremely hierarchical and/or large loops. To mimic this constraint within the LA framework, we introduce an energy penalty $U_{\rm gen}$ to discourage the formation of highly nested loops, which keeps both the loop hierarchy as well as the loop size in check (see Fig. 2). The hierarchical complexity of a LA, or, equivalently, the maximum loop size, is characterized by its maximum generation number g [26], where g is defined as the number of links separating the farthest node (in terms of links) from the origin (see Ref. [27] for the definition and requirement for an origin node). To construct an appropriate constraint for g, we note that the characteristic backbone length of a randomly branched polymer is a measure of g. It has previously been shown that for a randomly branched, self-avoiding chain with annealed branch points, this length scales with the size as $N^{0.66}$ [23]. Clearly, forcing g to a fixed value $g' = N^{\beta}$ with $\beta < 0.66$ confines the structure, and the strength of confinement increases with decreasing β. This insight allows us to construct an energy penalty for g in the same spirit as that for R_g, as given by $U_{\rm gen} = k_B T \exp[\sigma_g(g - g')]$, where g is the maximum observed generation number of the existing LA, g' is a fixed cutoff generation number, and σ is a weighting factor. The factor $\sigma_g \equiv g'/z$ again weights the degree of confinement according to the LA size and lattice coordination number. It should be noted that the generation-number confinement also automatically leads to spatial confinement of the LAs; hence, this constraint is stronger than the spatial confinement constraint introduced earlier. As in the case of spatial confinement, we use the generation-number constraint in conjunction with the excluded volume constraint. Consequently, the total energy of the LAs is given $U = U_{\rm gen} + U_{\rm exc}$.

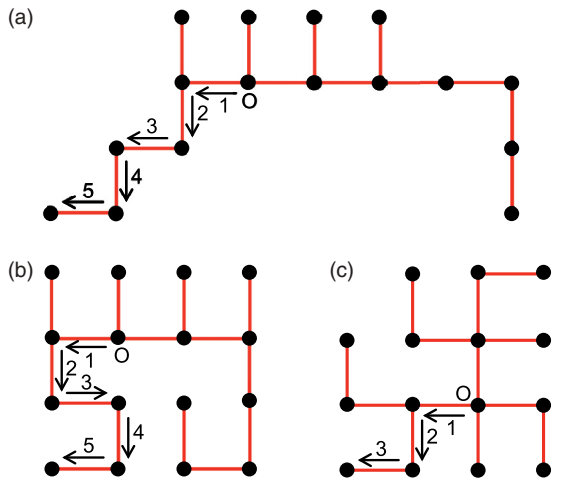

FIG. 2. (Color online) Schematic illustration of the difference between spatial and generation-number confinement. (a) Extended conformation of a 16-node LA obtained with only excluded volume constraints. (b) Spatial confinement of the LA can be achieved without explicit constraints on the hierarchy and loop size. In this specific example, we show how a 5-node structure emanating from the origin node (O) can be folded by a spatial confinement constraint. (c) Generation number confinement is achieved through explicit constraints on the hierarchy and loop size. In this specific example, we show how the 16-node LA can be restricted to a maximum generation number of 3 from the origin node by implementing such a constraint.

We employ Monte Carlo (MC) simulations to generate an ensemble of structures consistent with the four types of LA's. For each type, we examine five different chromosome sizes containing $N = 32$, 64, 128, 256, and 512 nodes, spanning a sufficiently broad range to allow reliable universal scalings to be derived. We explore two different strengths of spatial and generation-number confinements: moderate with $\alpha = 0.75$ and $\beta = 0.5$ and strong with $\alpha = 0.5$ and $\beta = 0.4$, though we report only results for the strong confinement unless stated otherwise. The initial N-node structure for the MC simulations is created by placing the first node, denoted as origin, at one of the lattice points. The remaining nodes are sequentially placed at unoccupied neighboring sites and connected via links to previously inserted nodes until the desired size N of the LA has been achieved. Each MC step involves an attempt to relocate an end node from one location to another. The node to be relocated is picked randomly from all nodes barring the origin node [27]; we denote the picked node by N_1. If N_1 is an internal node, the MC move is rejected and the

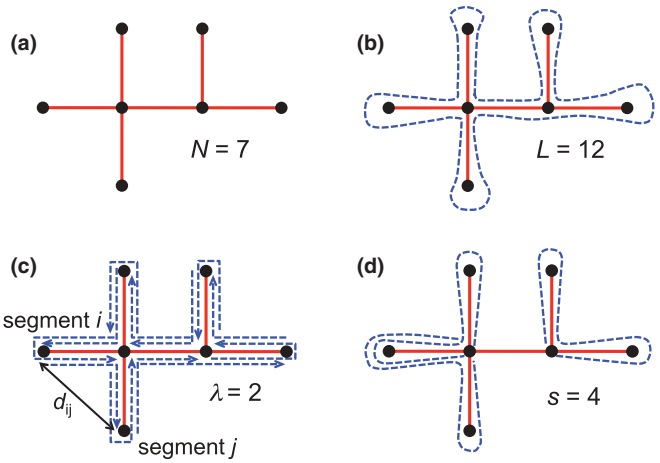

FIG. 3. (Color online) Enumeration of chromatin length, spatial distances, path lengths, and loop sizes in LA chromosomes. (a) For illustration, we picked a LA structure with $N = 7$ nodes. (b) The path of the chromatin fiber corresponding to this LA structure is depicted by the dashed blue line and yields a total contour length of $L = 12$. (c) Computation of spatial distances for segments separated by a path length $\lambda = 2$. All the 12 possible paths of length 2 are shown by dashed blue arrows. (d) Computation of number of loops of size $s = 4$. There are three such loops in the chosen LA structure, as indicated by the dashed blue curves.

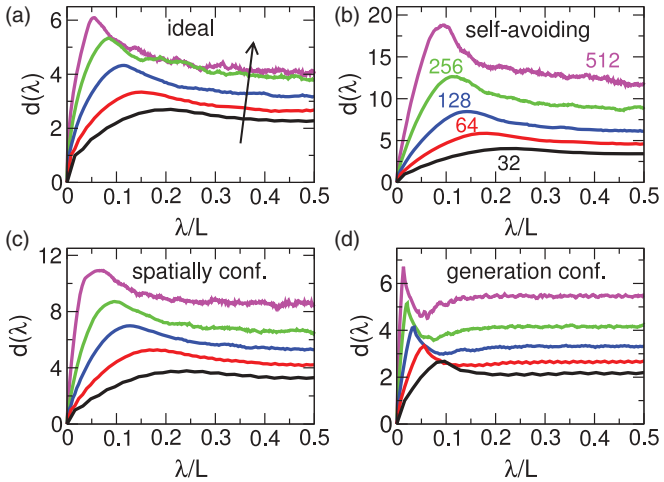

FIG. 4. (Color online) Root-mean-square spatial distance d as a function of scaled path length λ/L for (a) ideal, (b) self-avoiding, (c) spatially confined, and (d) generation-number confined LAs of sizes in the range $N = 32$–512. Arrow in (a) indicates the direction in which the plotted data shifts with increasing value of N, and it applies to all subfigures (a)–(d).

original structure is retained. If N_1 is an end node, we pick another node randomly from all nodes barring N_1 and the node attached to it; we denote the picked node by N_2. We then attempt to relocate N_1 to one of the $z = 6$ neighboring sites of N_2. This attempt is accepted with a Metropolis probability of $P_{\rm acc} = \min[1, \exp(-\Delta U/k_B T)]$, where ΔU is the energy difference between the proposed and original structures. The above procedure of relocating nodes to generate new structures satisfies the detailed balance condition. The simulations are performed for 5×10^7–10^8 MC steps.

III. RESULTS AND DISCUSSION

We first examine the *spatial* distance d between two LA segments as a function of their *genomic* separation λ along the chromatin for the four types of LAs introduced earlier. We note that the networklike structure of the LA [Fig. 3(a)] represents a chromosome whose underlying chromatin follows a path enveloping the network [Fig. 3(b)]. As indicated in Fig. 3(c), there exist $L = 2(N-1)$ pairs of segments (nodes) on this chromatin path that yield any given path length λ. These pairs of segments i and j with coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively, are separated by a spatial distance $d_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|$ and path length λ_{ij}. We are then left with determining the root-mean-square spatial distance $d(\lambda)$ for all pairs of segments i and j separated by path length $\lambda_{ij} = \lambda$ along the chromatin envelope. This calculation may be formally written down as

$$d(\lambda) = \left\langle \frac{1}{L} \sum_{i=1}^{L} \sum_{j=i}^{L} (d_{ij})^2 \delta(\lambda_{ij} - \lambda) \right\rangle^{1/2}, \qquad (1)$$

where the δ function ensures counting of distances for only the relevant pairs of segments and $\langle \cdots \rangle$ represents an ensemble average over all LA conformations generated during an MC simulation.

Figure 4 shows the spatial distances $d(\lambda)$ of the four types of LAs and five chromosome sizes N as a function of the scaled path length λ/L. All LAs exhibit an increase in d with λ with a scaling λ^ν for small path lengths. The exponent ν for the ideal LAs ($\nu = 0.47 \pm 0.1$) is consistent with that expected for random walk statistics ($\nu = 0.5$). The exponents for self-avoiding ($\nu = 0.64 \pm 0.05$), spatially confined ($\nu = 0.59 \pm 0.1$), and generation-number confined ($\nu = 0.67 \pm 0.07$) LAs are more consistent with self-avoiding random walk statistics ($\nu = 0.588$), though the generation-number confined show a small departure. Interestingly, for all LAs the rise in d peaks out at path lengths commensurate with the maximum generation number g of the LAs, or, equivalently, the spatial distances commensurate with the radius of gyration R_g. The sharpness of the peaks increases with the LA size. Similar peaks, albeit less pronounced, have been observed in FISH data; see, for example, plots in Refs. [8,9]. Beyond the maximum generation number g, a large number of paths adopt a near-constant distance from each other, forcing the distance scaling to be independent of the path length, similarly to the levelling off observed in FISH experiments [3,8]. Given that all four types of LAs exhibit such levelling off suggests that chromatin looping and *not* confinement might be the main contributor to the levelling off in spatial distances observed in FISH experiments.

We next obtain the distribution of loop sizes in our LA chromosomes. Given that each internal node represents a looping point, the probability P of finding a loop of size s therefore corresponds to the number of closed paths with $s/2$ links connected to each of the internal nodes in the LA [Fig. 3(d)] divided by the total number of loops of all possible sizes present in the LA. The ensemble-averaged $P(s)$ for the ideal, self-avoiding, spatially confined, and generation-number confined LAs are shown in Fig. 5. All four LAs show a power-law decrease in the probability of loop formation with

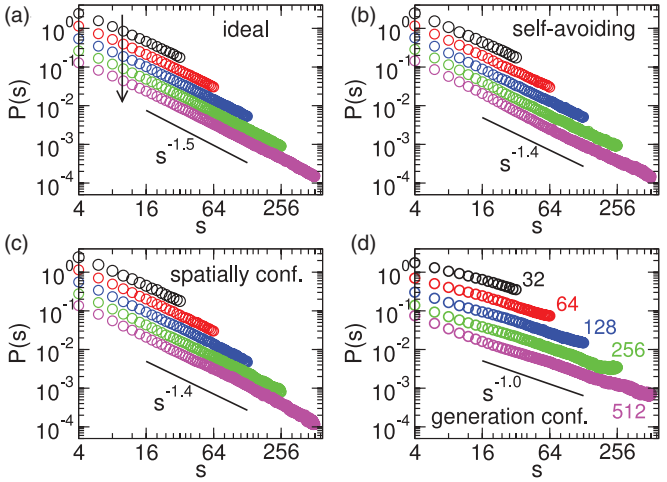

FIG. 5. (Color online) Looping probability $P(s)$ as a function of loop size s for (a) ideal, (b) self-avoiding, (c) spatially confined, and (d) generation-number LAs of varying sizes $N = 32$–512. For visual clarity, the distributions for $N = 32$, 64, 128, and 256 have been shifted upward by multiplying the original $P(s)$ by 16, 8, 4, and 2, respectively. The lines with slopes of -1.5, -1.4, and -1.0 are meant as guides to the eye. Arrow in (a) indicates the direction in which the plotted data shifts with increasing value of N, and it applies to all subfigures (a)–(d).

loop size (or path length) $P(s) \sim s^{-\chi}$, but show notable differences in the scaling exponent χ. The ideal LAs yield a scaling exponent of $\chi = 1.49 \pm 0.06$ in the asymptotic limit of $s >> 1$. This scaling agrees very well with the analytical result of $P(s) \sim s^{-1.5}$ [30] obtained for the probability of formation of structure of size $s/2$ in A_z condensation polymerization, where A_z represents a reactive monomer with z reactive sites. Note that z is analogous to the lattice coordination number in our simulations. When excluded volume interactions are introduced, $P(s)$ scaling is slightly lowered with $\chi = 1.38 \pm 0.03$. Further introduction of the spatial confinement constraint does not alter this scaling: $\chi = 1.41 \pm 0.05$. On the other hand, the introduction of a generation-number constraint reduces the scaling drastically to $\chi = 0.95 \pm 0.05$, which is in agreement with the exponent $\chi = 1.08$ observed in recent 3C experiments [15]. The generation-number confinement thus suppresses the formation of elongated structures dominated by a few large loops and many small ones, observed in self-avoiding LAs, and, instead, promotes the formation of a range of medium-sized loops, leading to the slower decaying $P(s)$ in confined LAs. Our results suggest that yeast genomes, which do not feel strong confinement effects as they are much more loosely packed than human genomes [20], might exhibit loop size scalings closer to $s^{-1.4}$ or $s^{-1.5}$ of ideal and self-avoiding LAs, respectively, rather than s^{-1} of generation-number confined LAs. Future 3C experiments on yeast genomes could test the validity of this hypothesis.

The above results demonstrate that the generation-number confined LA model captures best the observations from FISH and 3C experiments on human chromosomes. Specifically, the presence of random loops at all length scales, intrinsic to all four LAs, seems to be sufficient to capture the levelling off in spatial distances observed in FISH measurements. However,

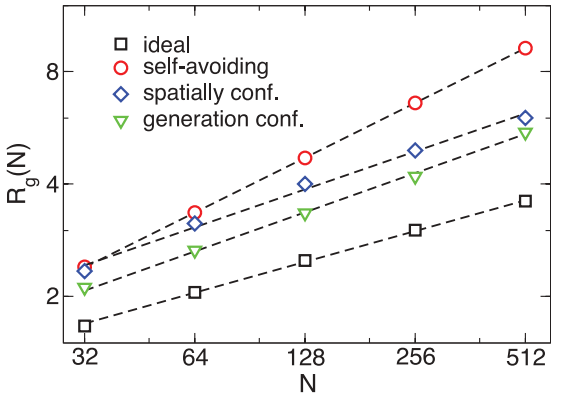

FIG. 6. (Color online) Radius of gyration R_g of ideal (black squares), self-avoiding (red circles), spatially confined (blue diamonds), and generation-number confined (green triangles) LAs for varying sizes $N = 32$–512. Error bars on all computed R_g are smaller than the size of the symbols.

additional constraints of excluded volume interactions along with suppression of elongated, hierarchical loops seem to be required for yielding the correct loop size distributions $P(s)$ observed in 3C experiments. Next, we investigate other structural properties of LAs to assess additional manifestations of the generation-number constraint.

In a system composed of loops, the radius of gyration R_g provides a more appropriate measure, as compared to d, for determining the scaling exponent ν [22], which is computed via the standard expression:

$$R_g = \left\langle \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (d_{ij})^2 \right\rangle^{1/2} \qquad (2)$$

where d_{ij} is the spatial distance between two nodes i and j. The ensemble-averaged R_g for the LAs are plotted in Fig. 6 as a function of size N. All LAs exhibit a power-law scaling $R_g \sim N^{\nu}$ with a constant exponent ν. The scaling exponent for the ideal ($\nu = 0.28$) and self-avoiding ($\nu = 0.49$) LAs compare well with the analytical results of $\nu = 0.25$ [28] and $\nu = 0.5$ [29] from field theory. The minor differences between the simulation and analytical results are due to finite-size effects. For spatially confined and generation-number confined LAs, ν lies between those of the ideal and self-avoiding LAs, and their magnitude depends on the extent of confinement. In particular, $\nu = 0.40$ for moderate confinement strengths of $\alpha = 0.75$ and $\beta = 0.5$ (results not shown), which gets further reduced to $\nu = 0.34$ and $\nu = 0.35$ for the strong confinement strengths of $\alpha = 0.5$ and $\beta = 0.4$, respectively. The latter ν values are similar to that obtained from the fractal globule model ($\nu = 1/3$), suggesting that both types of confinements cause the LAs to adopt a collapsed globulelike state [15].

Finally, we examine the internal organization of the LAs. The nodes can intrinsically be differentiated based on the number of segments n_b they are connected to, which is also indicative of looping activity. In Fig. 7, we compare the distance distribution from the center of mass for nodes possessing $n_b = 2$ and 5 neighbors. While the distributions for the two types of nodes are only marginally different in self-avoiding and spatially confined LAs, they show distinct

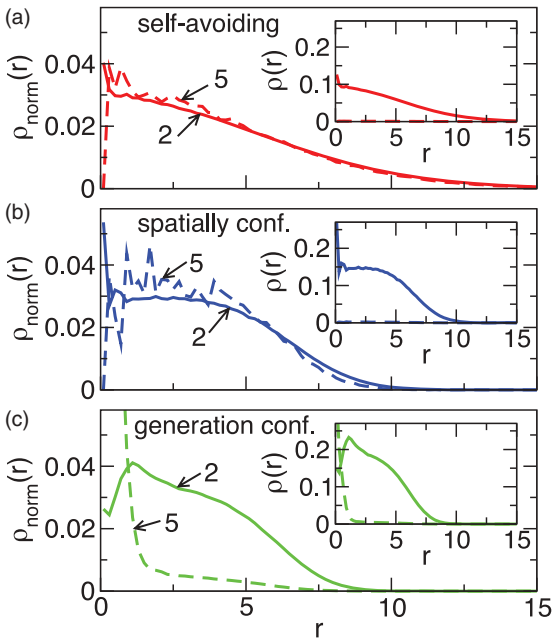

FIG. 7. (Color online) Normalized density of two-neighbor (solid lines) and five-neighbor (dashed lines) nodes as a function of distance from the center of mass for (a) self-avoiding, (b) spatially confined, and (c) generation-number confined LAs of size $N = 512$, respectively. The inset shows the corresponding absolute (unnormalized) densities for the three LAs.

segregation in generation-number confined LAs. Specifically, the active, five-neighbor nodes migrate toward the center due to confinement and the less active, two-neighbor nodes migrate toward the periphery. Thus, simple physical forces, such as those arising from excluded volume interactions, loop entropy, and confinement have the potential to yield complex functional patterns in chromosomes. Whether such a segregation into moderately and strongly looped chromatin domains occurs within eukaryotic chromosomes remains to be determined. Similarly, it would be interesting to examine whether other forms of functional segregation observed in chromosomes, such as migration of actively transcribed genes toward the nuclei centers [31] and formation of transcription, replication, and recombination "factories" [25,32,33], have any physical origin.

IV. CONCLUSION

The generation-number confined lattice animal model described above, accounting for both excluded volume and confinement constraints, successfully captures the observations from both the FISH and 3C experiments on human chromosomes. The presence of random sized loops at different hierarchies in this model ensures that the spatial distances between DNA segments level off at large separations. The excluded volume constraints invoked in the model act like topological constraints that prevent loops and looping points from interpenetrating. The generation-number confinement constraint imposed in the model is required for flattening the distribution of loop sizes and compressing the chromosomes into a collapsed state. The hierarchical organization of loops-within-loops built into the model ensures that the looping probability decays in a power-law manner over a range of loop sizes. The abstract lattice animal representation introduced here thus combines the advantageous features of the loop- and fractal-based representations into one common framework. Apart from reproducing FISH and 3C measurements, the model also predicts a collapsed-globule state of chromosomes and functional segregation of chromosome into weakly and strongly looped domains. Due to its abstract nature, the model is not amenable to studying detailed features of chromosomal organization that require a mapping of the lattice animal structure to specific features of DNA and higher-order structures of chromatin, such as predicting specific interactions between different parts of a given chromosome. However, this abstraction is useful for capturing global aspects of chromosomal organization, such as those discussed above, especially under conditions when the nature of the underlying higher-order structure of chromatin and the mechanisms leading to its organization remain unknown.

ACKNOWLEDGMENTS

We thank UCSD for funding this work. We also thank Cornelis Murre, Dario Meluzzi, and Jinting Liu for discussions.

[1] G. Felsenfeld and M. Groudine, Nature **421**, 448 (2003).
[2] J. Dekker, Science **319**, 1793 (2008).
[3] S. Jhunjhunwala, M. C.v an Zelm, M. M. Peak, S. Cutchin, R. Riblet, J. J. M. van Dongen, F. G. Grosveld, T. A. Knoch, and C. Murre, Cell **133**, 265 (2008).
[4] T. Cremer and C. Cremer, Nat. Rev. Genet. **2**, 292 (2001).
[5] L. A. Parada and T. Misteli, Trends Cell Biol. **12**, 425 (2002).
[6] T. Cremer, M. Cremer, S. Dietzel, S. Müller, I. Solovei, and S. Fakan, Curr. Opin. Cell. Biol. **18**, 307 (2006).
[7] R. K. Sachs, G. van den Engh, B. Trask, H. Yokota, and J. E. Hearst, Proc. Natl. Acad. Sci. USA **92**, 2710 (1995).
[8] J. Mateos-Langerak, M. Bohn, W. de Leeuw, O. Giromus, E. M. M. Manders, P. J. Verschure, M. H. Indemans, H. J. Gierman, D. W. Heermann, and R. van Driel, Proc. Natl. Acad. Sci. USA **106**, 3812 (2009).
[9] G. van den Engh, R. K. Sachs, and B. J. Trask, Science **257**, 1410 (1992).
[10] P. Hahnfeldt, J. E. Hearst, D. J. Brenner, R. K. Sachs, and L. R. Hlatky, Proc. Natl. Acad. Sci. USA **90**, 7854 (1993).
[11] J. Ostashevsky and C. Lange, J. Biomol. Struct. Dyn. **11**, 813 (1994).
[12] C. Münkel and J. Langowski, Phys. Rev. E **57**, 5888 (1998).
[13] M. Bohn, D. W. Heermann, and R. van Driel, Phys. Rev. E **76**, 051805 (2007).
[14] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, Science **295**, 1306 (2002).
[15] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A.

Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker, Science **326**, 289 (2009).
[16] D. V. Lebedev, M. V. Filatov, A. I. Kuklin, A. K. Islamov, E. Kentzinger, R. Pantina, B. P. Toperverg, and V. V. Isaev-Ivanov, FEBS Lett. **579**, 1465 (2005).
[17] A. Bancaud, S. Huet, N. Daigle, J. Mozziconacci, and J. Ellenberg, EMBO J. **28**, 3785 (2009).
[18] M. Bohn, and D. W. Heermann, PLoS ONE **5**, e12218 (2010).
[19] A. Rosa, N. B. Becker, and R. Everaers, Biophys. J. **98**, 2410 (2010).
[20] A. Rosa and R. Everaers, PLOS Comput. Biol. **4**, e1000153 (2008).
[21] M. Bohn and D. W. Heermann, J. Chem. Phys. **132**, 044904 (2010).
[22] B. V. S. Iyer, M. Kenward, and G. Arya, BMC Biophys. **4**, 8 (2011).
[23] A. M. Gutin, A. Y. Grosberg, and E. I. Shakhnovich, Macromolecules **26**, 1293 (1993).
[24] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, Europhys. Lett. **23**, 373 (1993).
[25] P. R. Cook and D. Marenduzzo, J. Cell. Biol. **186**, 825 (2009).
[26] A. R. Khokhlov and S. K. Nechaev, Phys. Lett. A **112**, 156 (1985).
[27] Typically, each node is considered in turn as the origin for the computation of the largest generation number of a LA. However, since the distribution of nodes with reference to any node should be equivalent for an exhaustively sampled structure, we find it sufficient to fix one node permanently as the origin node. Our choice of keeping the origin node fixed facilitates rapid computation of the maximum generation numbers while still yielding the correct ensemble of structures.
[28] T. C. Lubensky and J. Isaacson, Phys. Rev. A **20**, 2130 (1979).
[29] G. Parisi and N. Sourlas, Phys. Rev. Lett. **46**, 871 (1981).
[30] M. Rubinstein and R. H. Colby, *Polymer Physics* (Oxford University Press, New York, 2003).
[31] T. Misteli, Cell **128**, 787 (2007).
[32] P. R. Cook, Science **284**, 1790 (1999).
[33] J. S. Lucas, C. Bossen, and C. Murre, Curr. Opin. Cell. Biol. **23**, 318 (2011).