Chromosome structure predicted by a polymer model

Christian Münkel* and Jörg Langowski†
Biophysics of Macromolecules, German Cancer Research Centre (DKFZ), Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany and Interdisciplinary Center for Scientific Computing, University of Heidelberg, Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany
(Received 26 November 1997)

Human chromosomes were simulated by a polymer model in the presence of an excluded-volume interaction. The polymer chain of each chromosome was arranged into loops and several consecutive loops formed subcompartments. We observed the formation of distinct chromosome territories with separated chromosome arms and subcompartments in agreement with recent experiments. Mean spatial distances between markers agree with measurements under different preparation conditions. Different scaling properties were found depending on the genomic distance regarded. The radial density, the position of genes, and the overlap of chromosome arms and subcompartments agree with experimental results. The parallelization of the Monte Carlo and Brownian dynamics algorithms and a method for the computation of confocal images is described, which allows a direct comparison with results from confocal light microscopy. [S1063-651X(98)12104-9]

PACS number(s): 87.15.−v, 36.20.−r, 81.05.Ys, 07.05.Tp

I. INTRODUCTION

The structure and dynamics of chromatin is known to play an important role in eukaryotic gene expression and differentiation [1]. However, in spite of the rapid progress in the physical mapping of the human genome, the higher-order three-dimensional arrangement of the DNA in the cell nucleus is still largely unknown. Even the structure of the chromatin fiber is still discussed (for reviews see [2]).

Approaches that have been successful for the analysis of other cell compartments have failed to give us detailed information about the cell nucleus. Classical electron microscopy techniques identified some nuclear compartments (for a review see [3]). Only recently, careful electron tomography studies in combination with light microscopy (e.g., [4]) indicated undisturbed higher-order chromatin structures. *In situ* hybridization of mammalian and plant cells demonstrated that the DNA of a chromosome is not distributed throughout the entire nucleus but limited to a territory, i.e., a subvolume of the nuclear space (e.g., [5–8]). However, experiments using light microscopy were limited by the resolution of several hundred nanometers, i.e., about $\frac{1}{20}$ of a territory diameter. Therefore, the spatial organization of the DNA *within* a territory remained unclear. Proposed models range from highly randomly intermingled fibers [9–13] to certain levels of organization [14–17] or focused on the functional organization [18,19]. These different models can be evaluated by a comparison of experimental data to computer simulations based on polymer models, which link our knowledge on small scales (e.g., the DNA sequence) to observed higher-order three-dimensional structures.

The ''multiloop subcompartment'' (MLS) model presented in this work predicts structural flexibility together with a high degree of compartmentalization of chromosome territories. The parallelization method, which is described here for Monte Carlo and Brownian dynamics, is applicable for other polymer or many-particle systems with short-range interactions as well.

II. POLYMER MODEL AND SIMULATION TECHNIQUES

A. Polymer model

Each chromosome in a cell nucleus consists of a single DNA fiber [20]. We assume that all chromosomal DNA can be described in terms of the next higher folding motif: the chromatin fiber. The detailed structure of chromatin fibers in interphase nuclei is still discussed [2], but its average properties, like those of any linear chain molecule, can be described by polymer physics.

We represent the chromatin fiber of a chromosome by a linear chain of segments (Fig. 1). Each segment has an assigned stretching energy

$$U_s(l)=\frac{k_BT}{2\delta^2}(l-l_0)^2 \quad \text{(stretching)}, \tag{1}$$

where l is the actual length of the segment and l_0 its central length. An absolute temperature $T=300$ K and Boltzmann's constant k_B were used. The stretching rigidity of chromatin is not known; however, we have to assume some stretching elasticity to keep the Brownian dynamics simulations numerically stable without the introduction of additional constraints. We used $\delta=0.1$, which allows fluctuations of about 10% of l_0. For the Brownian, molecular, and hybrid molecular-dynamics simulations the stretching force at bead i was computed as

$$\boldsymbol{F}_i^s=-\frac{k_BT}{\delta^2}[(l_{i-1}-l_0)\boldsymbol{u}_{i-1}-(l_i-l_0)\boldsymbol{u}_i] \quad \text{(stretching)}, \tag{2}$$

where u_i indicates the unit vector of segment i.

To keep the number of segments as small as possible the rigidity of the chromatin fiber was modeled by setting the

*Electronic address: C.Muenkel@DKFZ-Heidelberg.de
†Electronic address: Joerg.Langowski@DKFZ-Heidelberg.de

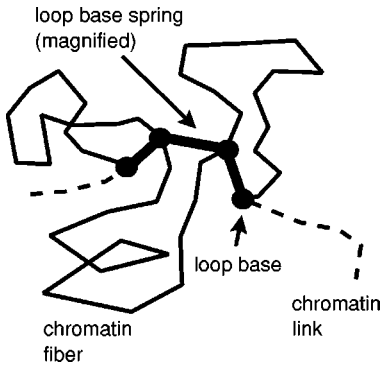

FIG. 1. Schematic drawing of the multiloop subcompartment model with subcompartments consisting of about 120 kbp sized loops (with a contour length of approximately 1.2 μm), which are formed by a harmonic potential between the loop bases. About ten loops form a subcompartment (approximately 1 Mbp). Subcompartments are linked by small chromatin fibers.

segment length l_0 equal to the Kuhn length L_K. For a random walk with a density of d base pairs per nanometer contour length the Kuhn segment length L_K can be estimated by the ratio of the mean distance between probes R and the genomic distance x between the markers in base pairs:

$$R^2 = \frac{L_K}{d} x. \tag{3}$$

Using measured distances below 1 mega base pairs (Mbp) of van den Engh *et al.* [12], one obtains an initial slope L_k/d of about 2.5 ± 0.5 μm^2/Mbp. A chromatin density of about 6 nucleosomes per 11 nm ($d \approx 105$ bp/nm) [21] results in $L_K \approx 260$ nm, in agreement with earlier estimates ($200-359$ nm [11]). We used $L_K = 300$ nm (about 30 kbp) for the simulations. For a more detailed local description of the fiber we substituted each segment by n smaller ones with length $b_0' = L_K/n$ ($n=6$ was used throughout this work). The same rigidity can be achieved by a harmonic bending potential

$$U_b(\beta) = \frac{k_B T}{2\psi^2}\beta^2 \quad \text{(bending)}, \tag{4}$$

which acts between two adjacent segments. β is the angle between those two segments and ψ is related to the Kuhn length L_K by $L_K = b_0/2\psi^2$ approximately. The corresponding bending force is

$$\boldsymbol{F}_i^b = -\frac{k_B T}{\psi^2}(\boldsymbol{A}_i - \boldsymbol{A}_{i-1} + \boldsymbol{B}_i - \boldsymbol{B}_{i-1}) \quad \text{(bending)}, \tag{5}$$

with

$$\boldsymbol{A}_i = \beta_i \frac{\boldsymbol{u}_{i+1} - \boldsymbol{u}_i \cos\beta_i}{b_i \sin\beta_i}$$

$$\boldsymbol{B}_i = \beta_{i-1} \frac{\boldsymbol{u}_{i-1} - \boldsymbol{u}_i \cos\beta_{i-1}}{b_i \sin\beta_{i-1}}. \tag{6}$$

Because no experimental data about the torsional interaction of the chromatin fiber are available and certain proteins can relax torsional stress, we did not include any in our simulations.

Chromosomes and even typical distances between parts of the chromatin fiber are much larger than the van der Waals interaction range or the electrostatic Debye length for the ionic conditions present in cell nuclei. Therefore, we approximated these interactions by a short-range excluded-volume interaction

$$U_e(r) = U_e^0 k_B T\left(1 + \frac{r^4 - 2r_c^2 r^2}{r_c^4}\right) \quad \text{(excluded volume)} \tag{7}$$

between all beads separated by a distance r in the interval $[0.0, r_c]$. r_c ensures a spatial separation of all parts of the chromatin fiber and was chosen 50% larger than the radius of the chromatin fiber. Because of the finite barrier at $r=0$ and the vanishing forces

$$\boldsymbol{F}_i^e(\boldsymbol{r}) = -U_e^0 \frac{4k_B T}{{r_c}^4}(r^2 - {r_c}^2)\boldsymbol{r} \quad \text{(excluded volume)} \tag{8}$$

at the boundaries $r=0$ and $r=r_c$ this interaction models the repulsion as well as a certain probability of crossing of chain segments. It has been argued [22] that such a chain crossing is essential to enable disentanglement of replicated chromosomes; this chain crossing could be mediated by topoisomerase II. Most of the simulations were performed with a barrier $U_e^0 = 0.1 k_B T$. Frequent crossings were observed during the initial decondensation of the chromosomes. The equilibrium properties presented below, however, were hardly affected by using the ten-fold higher barrier $U_e^0 = 1 k_B T$. Other properties, e.g., the number of knots, may depend on the height of the barrier.

The presence of adjacent chromosome territories is approximated by an additional energy barrier $0.1 U_e^0$ (equal to the intraterritorial crossing barrier) for each segment outside the territory, which was approximated by a sphere. The territory radius $R_{\text{territory}}$ was computed by the nuclear radius R_{nucleus} times the cube root of the ratio of chromosome DNA content $c_{\text{chromosome}}$ and genome size c_{genome}:

$$R_{\text{territory}} = R_{\text{nucleus}}\left(\frac{c_{\text{chromosome}}}{c_{\text{genome}}}\right)^{1/3}, \tag{9}$$

a relation confirmed experimentally [23].

The average end-to-end distance of a free chromatin chain can be estimated for, e.g., chromosome 1 as $\sqrt{\langle R^2 \rangle} = (b_0/d)x \approx \sqrt{300 \text{ nm}/(105 \text{ bp/nm}) \times 2.5 \times 10^8 \text{ bp}} \approx 27$ μm and exceeds the size of three-dimensional (3D) reconstructed chromosome territories by far [24]. Therefore, additional topological constraints need to be enforced. Such constraints had been introduced, e.g., as a rigid territory boundary [10] or large loops in the Mbp range [11]. Recently measured interphase distances along a whole chromosome showed a crossover from the initial random walk to a more compact structure [25] and were interpreted as an additional backbone polymer, which folds the fiber into giant loops [13].

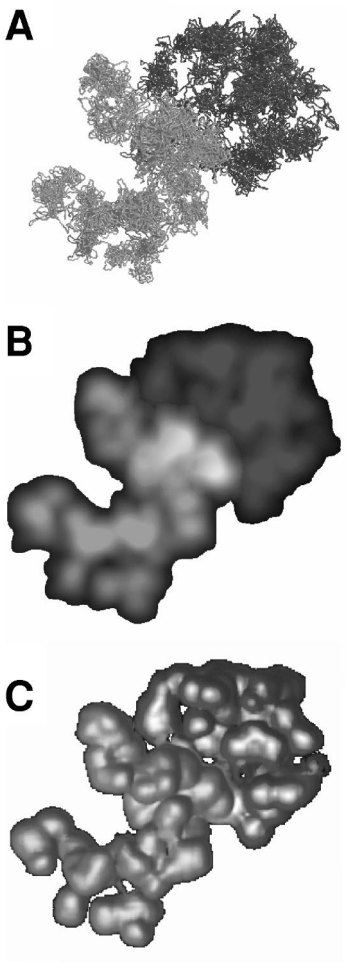

FIG. 2. Three-dimensional structure of a simulated chromosome 3 with both arms shaded in different levels of gray. (a) Polymer segments visualized by cylinders with the diameter of the chromatin fiber (30 nm), (b) Computed central section corresponding to a confocal laser scanning microscope (250-nm lateral and 750-nm axial resolution). (c) Three-dimensional volume rendering based on all subsequent confocal sections.

However, no such giant loops in the Mbp range were found by biochemical experiments [16,21], fluorescence *in situ* hybridization [18], metaphase spreads [17,26,27], nuclear matrix preparations [28], and electron tomography studies [4]. In contrast, indications for loop sizes of about 100 kbp were found [29]. We used small loops of about 120 kbp (four Kuhn segments). The springs between the loop bases in Fig. 1 were modeled similar to the stretching potential [Eqs. (1) and (2)] but with a vanishing minimum energy length l_0,

$$U_{\text{base}}(l)=\frac{k_BT}{2\delta^2}l^2 \quad \text{(stretching)}. \tag{10}$$

About ten such loops form a subcompartment as shown in Figs. 1 and 2(a). The exact number of loops in each subcompartment was derived from the observed band pattern. For lack of sufficient experimental data we extrapolated the band pattern of the standard human 850 band ideogram [30] to the earliest stage of human prophase (about 3000 bands [31]) by subdividing each standard band into three equally sized parts. This results in a band size of about 1.5 Mbp. The number of loops in each subcompartment is computed by assuming that the amount of DNA in each band is proportional to the band size in the ideogram and subdividing this number by the size of a loop. Additionally, we simulated a structural modification of the MLS model where several successive subcompartments are opened and their chromatin forms a single giant-loop domain in the Mbp range. Except for this different boundary condition, the same interactions and simulation techniques were used. In contrast to the analytical random-walk–giant-loop (RW-GL) model [13], where the number or size of the loops was adapted to fit measured interphase distances, all comparisons of the MLS model with experimental data were performed without any further adjustments of the parameters noted above. No additional backbone polymer is assumed in the MLS model.

B. Simulation algorithms

The Brownian dynamics simulations were performed according to the Ermak-McCammon algorithm [32]. Hydrodynamic interactions between segments can be neglected because of hydrodynamic shielding [33] and were not computed during the simulation. For the Monte Carlo simulations several moves were used at each step. Local changes were performed by a rotation of beads around the axis given by both adjacent neighbor beads. Concurrent updates of adjacent beads were avoided. Because this move preserves the length of all segments the time scale of our simulations was not limited by the fast bond vibrations. Global rearrangements were accelerated by pivot moves [34]. One bead of the chain and a vector on the unit sphere were chosen randomly. Then all beads below (or above) the chosen bead are rotated by a random angle around that vector. While both moves are well suited for simulations of linear chains, they change the orientation of our loops within the subcompartments only slowly. Hence we introduced an additional loop-rotation move, where a whole loop is rotated randomly around the loop base in the same way pivot moves are performed. In the case of spatially different loop bases the loop rotation move will separate them further. We corrected for this by a steepest-descent relaxation step, where all beads are moved according to the force acting on them. The stability of the algorithm in the presence of loop rotation and relaxation moves was checked for some 10×10^6 Monte Carlo steps in total.

C. Parallelization

Because of the large numbers of segments the computational complexity of the pairwise excluded-volume interaction (7) dominates the computation time. All other interactions used (e.g., stretching and bending) scale linearly with the number of segments N only. For short-range interactions the sum over all pairs

$$U_e^{ap}=\sum_{j=2}^{N}\sum_{i=1}^{j-1}U_e(r_{ij}) \quad \text{(pairwise)} \tag{11}$$

can be avoided by using the linked-cell algorithm (see, e.g., [35]), where the three-dimensional space is subdivided into

cubes. The side length of the cubes is equal to the pairwise interaction range. Only interactions with beads within the same cell and one-half of all 26 neighbor cells have to be regarded (n_k is the number of beads in cell k):

$$U_e^{lc}=\sum_{k}^{\text{cells}}\left\{\sum_{i}^{n_k}\sum_{j}^{i}U_e(r_{ij})+\sum_{l}^{13}\sum_{i}^{n_k}\sum_{j}^{n_l}U_e(r_{ij})\right\}\quad\text{(linked cells)}\qquad(12)$$

Accordingly, the number of computations scales linearly with N.

Various regular communication schemes were proposed for uniform mappings of cells to processors in the context of molecular dynamics (for reviews see [36]), but they generally scale for large number of particles and an uniform particle distribution only. We avoided the communication overhead by using asynchronous, buffered communication between processors, which already worked well in the context of membranes [37]. At the beginning of a Brownian dynamics step all beads belonging to a processor are stored in the first of two lists. After computing the interactions for all beads in a cell with the 26 neighboring cells the equations of motion are integrated and the updated bead positions (and other data if necessary) are communicated to the neighboring processors asynchronously. During the computation each processors receives the updated beads and stores them in a second list together with the beads, which remain on the processor. When all computations of a Brownian dynamics step are completed the beads in the second list are sorted into the first list according to their positions in cell space. As an additional advantage this methods maps the spatial locality of the beads on a locality in memory. This leads to an improved computation speed on cache-base computers.

The strongly irregular spatial distribution of beads limits the parallelization efficiency in the case of a regular mapping of cells to processors. A better load distribution can be achieved by a dynamic load balancing algorithm based on a Monte Carlo approach. We used a dynamic mapping of the three-dimensional cell space onto a two-dimensional processor grid, i.e., all cells with the same x- and y-coordinates are mapped on the same processor. Initially, the mapping is a regular two-dimensional grid. An improved load balancing is achieved by moving the vertices of the grid according to a ''pressure'' that is induced by an imbalance in the number of beads in the four domains around the vertex. If a random move of a vertex leads to an increase in the cost function it may be rejected according to a Metropolis test. In general, this algorithm leads to a distortion of the grid, which also decreases the performance: While domains with a high load are shrunk this way, they tend to develop spikes. We therefore introduced an additional cost term, which is proportional to the angle between each two adjacent grid edges. In the presence of this additional cost the area of high load domains still decreased, but their shape remained nearly quadratic.

The simulation software was designed according to an object-oriented approach and is coded in the language C++ (about 15.000 lines). Computer-system-specific communication details are hidden in a *communication object*, which uses standard message passing library functions. The results presented below were mostly computed by the parallel Monte Carlo algorithm and took about 800 days of single CPU time on a Parsytec PowerXplorer with 4 and a PowerGC with 192 PowerPC 601 (80-MHz) processors.

D. Simulation of confocal images

For the visual exploration of the three-dimensional chromosome organization we developed the *virtual confocal microscope*, which allows a direct comparison with microscopical images. Figure 2(b) shows a computed central section of the configuration in Fig. 2(a). Pointlike fluorescent markers were positioned on the chromatin fiber with a separation smaller than the diameter of the fiber. The imaging properties of a confocal microscope were modeled by an image convolution with a Gaussian point spread function. Image operations were performed on a three-dimensional 256^3 grid of 50-nm^3 voxels. A lateral resolution of 250 nm and an axial resolution of 750 nm was assumed for the visualizations, which is similar to that of the confocal light scanning microscopes used in many experiments. Additionally, we rendered three-dimensional reconstructions by volume rendering (which is similar to the ray-tracing method) based on the computed sections. Our rendering software is based on the public domain volume rendering library VOLPACK [38].

E. Computation of domain overlap

The organization of DNA within a chromosome territory is studied by the overlap of chromosome arms and subcompartments. In the experiment, for example, both arms are painted with different fluorescent markers, which are detected in two different color channels. The relative overlap O can be computed by comparing the intensities r and g of the red and green channels directly, but generally the results will be strongly influenced by the background intensities. Summing over all voxels of the three-dimensional array $[N_x,N_y,N_z]$,

$$O_{\text{simple}}=\frac{1}{N_xN_yN_z}\sum_{k_x,k_y,k_z}^{N_x,N_y,N_z}r_{k_x,k_y,k_z}g_{k_x,k_y,k_z},\qquad(13)$$

includes contributions from the randomly distributed background also. Probe and background signal correlations can be separated due to their different properties under relative pixel shifts between both color channels because correlations of an isotropic background are invariant. Unfortunately, averaging over all relative shifts leads to an increase of computational complexity by the third power of the linear image size. This can be avoided by computing the correlations in Fourier space [39]. After transforming the 3D images of both color channels by two 3D fast Fourier transformations (FFT's) the real-space correlation

$$\mathcal{C}(r,g)_{j_x,j_y,j_z}\equiv\sum_{k_x,k_y,k_z}^{N_x,N_y,N_z}r_{j_x+k_x,j_y+k_y,j_z+k_z}\;g_{k_x,k_y,k_z}\qquad(14)$$

is equivalent to a multiplication of the transformed intensities R and G in Fourier space

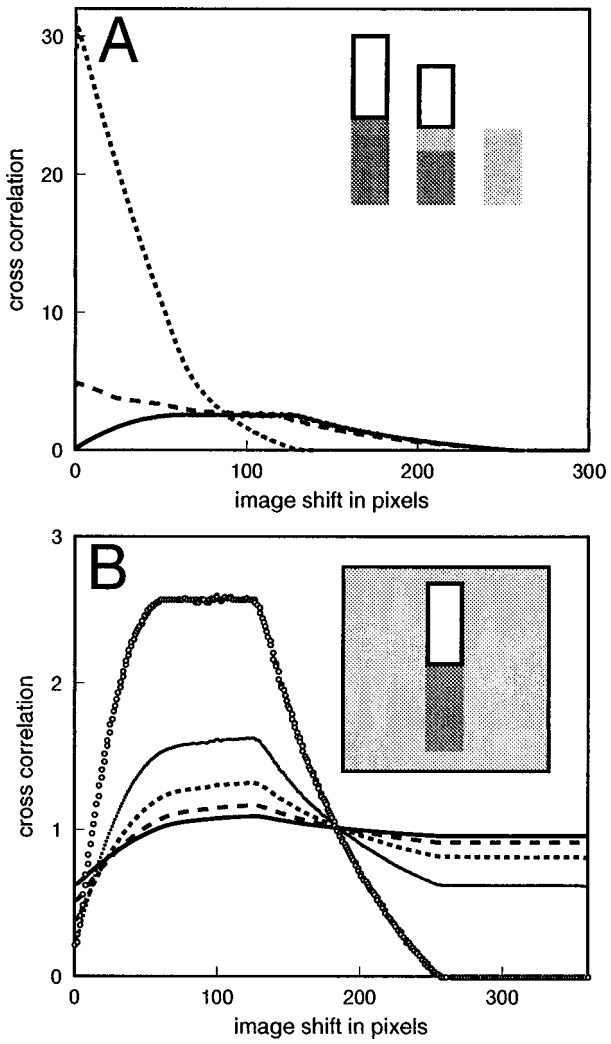

FIG. 3. Determination of spatial overlap by cross correlation of color channels. (a) Computed cross correlation dependent on the relative shift between the red and green planes for nonoverlapping domains (continuous line), a small overlapping region (long-dashed line), and a complete overlap of domains (short-dashed line) as depicted in the inset. Without background signals the inverse of the cross correlation is equal to the relative size of the overlapping area, which was 1/32 of the whole image for complete overlap. (b) In the presence of a randomly distributed background (10% of maximal signal intensity, Gaussian distribution) the cross correlation reaches a plateau of about 1 for large shifts (continuous line). The background contribution can be detected by increasing the threshold (dashed lines) until the plateau vanishes for large shifts (circles).

$$\mathcal{C}(r,g) \Leftrightarrow RG^* \tag{15}$$

and a subsequent retransformation of the correlation data.

Painting both arms of a chromosome entirely results in the three different classes of probe arrangements as shown in Fig. 3(a). Two nonoverlapping domains have a vanishing correlation in the absence of shifts. The correlation increases with shift distance until a plateau is reached. For large shifts the correlation vanishes again because the domains are much smaller than the image size and most shifts result in nonoverlapping domains. In the presence of an overlapping region a small initial peak is added to the correlation. Completely overlapping domains result in an decreasing correlation without any plateau. In the absence of background signals, the inverse value of the correlation is equal to the fraction of overlap area relative to the size of the image. In Fig. 3(a) the size of the completely overlapping area is $1/8 \times 1/4 = 1/32$ of the image size, in agreement with the correlation without shifts but complete overlap. A randomly distributed background signal leads to a correlation close to 1 [Fig. 3(b)]. By increasing the threshold the original correlation [Fig. 3(a)] is recovered and the optimal threshold can be determined by a vanishing correlation for large shifts.

III. RESULTS

A. Interphase distances

While the characteristic pattern of chromosome bands in metaphase indicates a certain degree of unique structure, measured spatial distances between markers vary strongly in interphase [12,25,40–42], suggesting many different possible foldings of the chromatin. However, this does not necessarily imply a randomly intermingled network of the chromatin fiber. The MLS model predicts a highly compartmentalized chromosome territory structure [Fig. 2(c)]. Nevertheless, predicted interphase distances agree well with the experimental data [Figs. 4(a) and 4(b)]. Chromosome 4 was simulated for 10^6 Monte Carlo steps. Projected interphase distances computed for pairs of markers within the region $4p16.3$ [Fig. 4(a)] are in good agreement with experiments [12], but require a local decondensation of about three subsequent subcompartments into a giant loop. This model is still different from the random-walk–giant-loop model [13] because we have the additional excluded-volume interaction and each two subsequent giant loops are separated by a MLS-model subcompartment consisting of small loops. We note that the experimental distance measurements [25] were performed on hypotonically swollen nuclei that were fixed with 3:1 methanol:acetic acid and air dried. Since this procedure leads to extensive flattening of the nuclear shape, Yokota *et al.* also performed distance measurements for markers in the same region in formalin fixed nuclei under conditions, which better preserved their three-dimensional nuclear structure. In these nuclei they measured considerably smaller distances, which are not compatible with the presence of giant loops but fit very well with the prediction of the MLS model [Fig. 4(a), dashed line]. It should be reemphasized that the loop size (120 kb, i.e., four segments of the polymer model) was not selected to fit these experimental interphase distances, but motivated by the literature [16,21,26,27]. The results are unsensitive to an increase of the loop size by a factor of 2 [240 kbp, eight segments; short-dashed line in Fig. 4(a)], which is much smaller than the subcompartment size. No agreement can be achieved for much larger sizes, i.e., for a factor of 40, which corresponds to giant loops of about 5 Mbp. This indicates that giant loops are created as a result of the preparation conditions used.

Although the presence or absence of giant loops result in different distances on the Mbp level [Fig. 4(a)], all models predict similar distances in the 200-Mbp range in agreement with the experimental data of both preparation techniques [Fig. 4(b)]. However, the MLS model predicts interphase chromosomes to be in a globular state, where spatial distances between markers in a genomic distance d scale as $d^{1/3}$

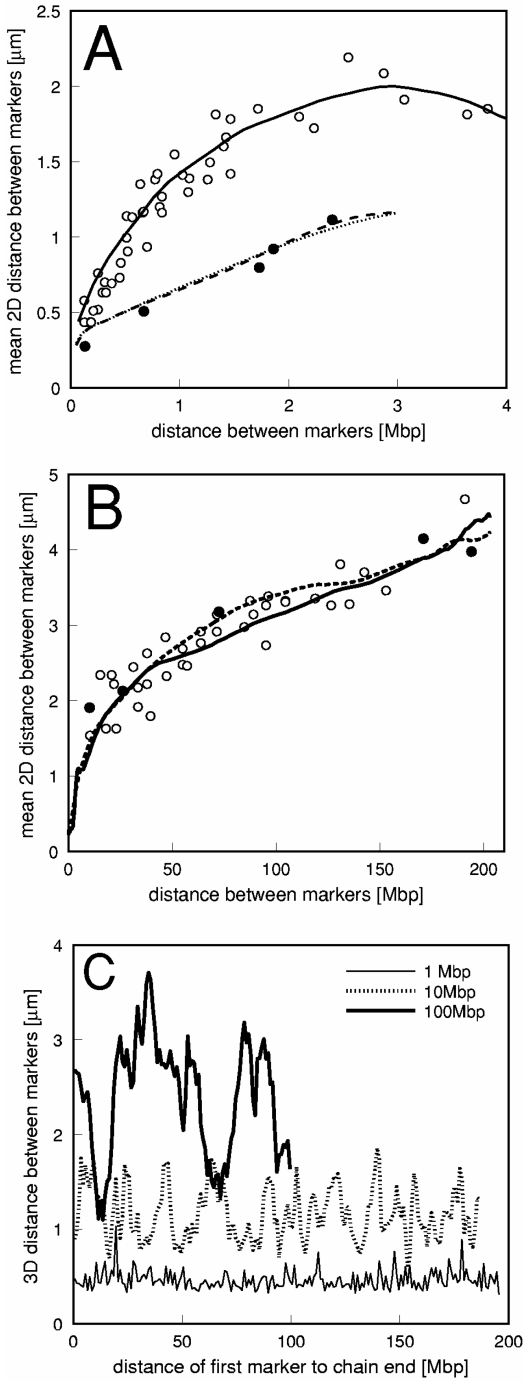

FIG. 4. (a) Mean projected interphase distances of markers in region $4p16.3$ of chromosome 4. Results for the multiloop subcompartment model (long-dashed line 120-kb loops; short-dashed line, 240-kb loops) agree with experimental results for nonhypotonically swollen nuclei (filled circles). A local rearrangement of about three subsequent subcompartments into a giant loop results in distances (continuous line) corresponding to experiments with hypotonically swollen nuclei (open circles). (b) The larger marker distances on chromosome 4 (open circles, hypotonically swollen, filled circles, nonhypotonically swollen nuclei) are less dependent on preparation conditions, in agreement with our model simulations with (continuous line) and without (dashed line) giant loops. (c) For a single MLS-model configuration distances vary strongly depending on marker distance, reflecting the variability of the folding pattern at the highest level.

[43], while the RW-GL model results in an exponent of 1/2. A double logarithmic plot of the experimental data from Yokota *et al.* [25] yields an exponent $0.32(2)$, in agreement with the MLS model (data not shown).

In any individual nucleus, however, three-dimensional distances between markers depend strongly on the position of the marker pair. Computed distances in Fig. 4(c) for markers separated by 1, 10, and 100 Mbp vary up to 3 times. While this flexibility allows interactions of distant genes on a chromosome, such interactions may be the cause for the different folding patterns observed in experiment [44].

B. Spatial chromatin distribution

The globular state characterized by a scaling $d^{1/3}$ corresponds to an approximately uniform average, radial chromatin density within the entire chromosome territory (neglecting differences in local chromatin condensation). In agreement with the distance measurements described above and experimental observations we find a plateau in the radial chromatin distribution for an average of 200 MLS-modeled chromosomes 3 over 10^6 Monte Carlo steps [Fig. 5(a), solid line]. The constant density is caused by the interplay between the excluded-volume interaction and the entropy of the fluctuating loops, which leads to adjacent but almost nonoverlapping subcompartments visible in Fig. 2(c), which themselves have an approximately uniform average density distribution. It is not caused by the straight initial metaphase configuration, which was used at the beginning of the simulation and could be preserved by a prohibited crossing of chromatin loops. The barrier $U_e^0 = 0.1 k_B T$ for chain crossing, which model topoisomerase II activity, allows frequent crossing of chains, which was indeed observed during the initial phase of the simulations. The arms, however, always remained separated essentially.

In contrast, the vanishing barrier in the RW-GL model leads to a majority of configurations where the chromosome is intermingled to a large extent. Averaging the radial density of 200 RW-GL-modeled chromosomes 3 over 10^6 Monte Carlo steps results in a decrease without plateau [Fig. 5(a), dashed line], which is not observed experimentally. The high central density is caused by many phantom (i.e., nonexcluding) segments separated less than the diameter of the chromatin fiber.

The structural difference between both models is therefore caused by the presence of the excluded-volume interaction and adjacent chromosome territories. We approximated these other chromosomes by a spherical territory boundary, but since this boundary had a low-energy barrier (see Sec. II A) the chromatin fiber was not limited to this sphere. The decay of the average density in the range 1500–3000 nm indicates fluctuations of the territory shape, which are also observed in experiments.

Because of the random folding of the chromosome one expects similar radial distributions for the parts of the chromosome corresponding to *R* and *G* domains. Indeed, Fig. 5(b) shows almost identical distributions for R and G-subcompartments and both models. Experimentally, the same similarity was found for early and late replication domains [45], which are assumed to be almost identical to *R* and *G* domains.

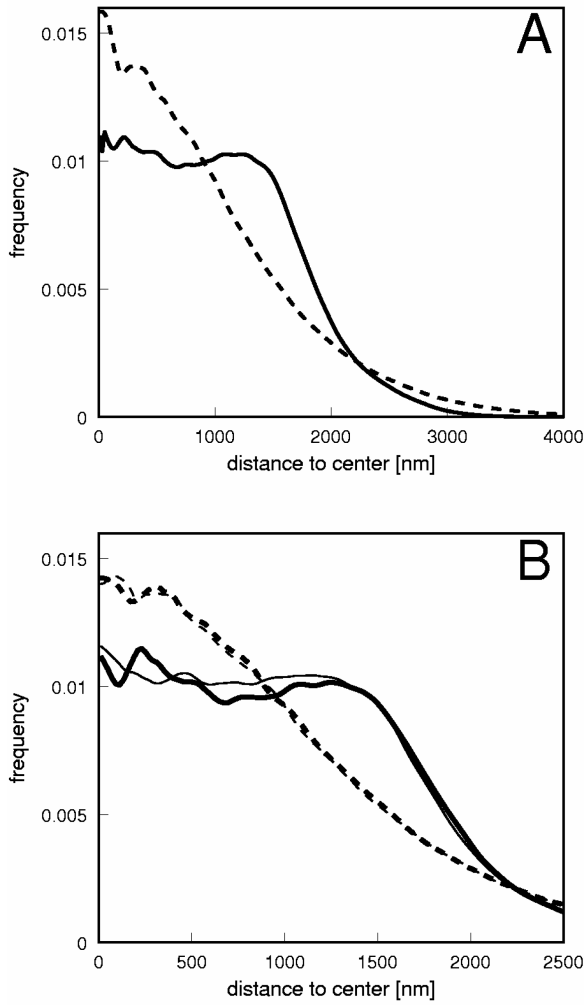

FIG. 5. (a) Average radial density of 200 MLS-modeled chromosomes 3 (continuous line) and RW-GL-modeled chromosomes (dashed line). The plateau reflects the compact territory structure predicted by the MLS model while the RW-GL model shows a typical random walk distribution with very high central density. (b) For both models the average radial densities for *R*-band domains (thick lines) and *G*-band domains (thin lines) agree within statistical errors.

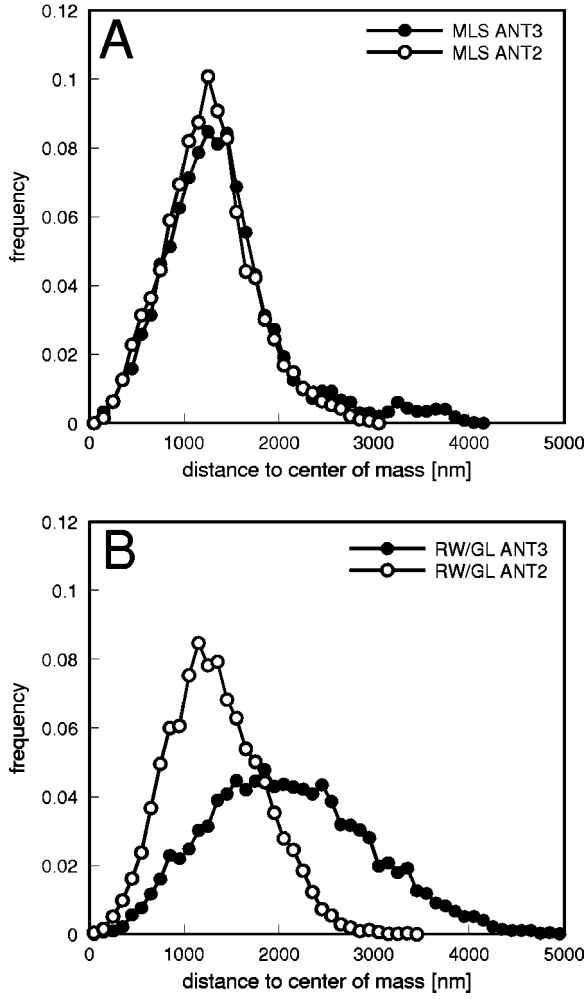

FIG. 6. Distances between genes ANT3 (filled circles) and ANT2 (open circles) and the chromosome center of mass. ANT3 is located close to the chain end and ANT2 is close to the chain center. (a) Only a small difference at large distances is visible for the MLS model. (b) Without excluded volume both distributions differ significantly for the random-walk–giant-loop model.

C. Position of genes

While both models do not predict different arrangements of *R* and *G* domains, the position of certain genes can be different. A specific association of chromatin to other nuclear domains, e.g., nucleoli or the nuclear lamina, may lead to a different average position of coding and noncoding sequences, which was observed experimentally [19]. Even more interesting is the relation between transcription (i.e., activity) of a gene and its position relative to the chromosome territory. This position is also be influenced by the geometric constraints imposed by the folding of the chromosome. We therefore simulated chromosome *X* according to the MLS and the RW-GL model and computed the distances between ANT2 and ANT3 genes and the territory center of mass. Gene ANT3 is located in the pseudo-autosomal region (i.e., close to one end) of the *X* chromosome and is being transcribed in both the active and inactive *X* chromosome. ANT2 is located near the center of the chromosome and is active in the active *X* chromosome only. The distance distributions for the MLS model in Fig. 6(a) differ only slightly. The RW-GL model, however, results in completely different distributions [Fig. 6(b)], i.e., the results depend strongly on the sequential position of a gene. A comparison of these predictions and of computed distances between genes and the territory surface with experimental data will also help to explain the mechanisms of *X* chromosome inactivation, which is accompanied by structural changes [24].

D. Overlap of chromosome arm and band domains

The results described above suggest that MLS-model chromosome arms are essentially separated, while the RW-GL chromosomes can be highly intermingled because of the lack of excluded-volume interactions. In agreement with this expectation and the examples in Fig. 3, the correlation functions in Fig. 7 of the MLS-model chromosomes show an initial increase and those of the RW-GL-model chromosomes a single peak. A higher simulated resolution of 100 nm^3 reduces the computed overlap and the correlation signal for small shifts increases by 20–40 % (Fig. 7, thin lines).

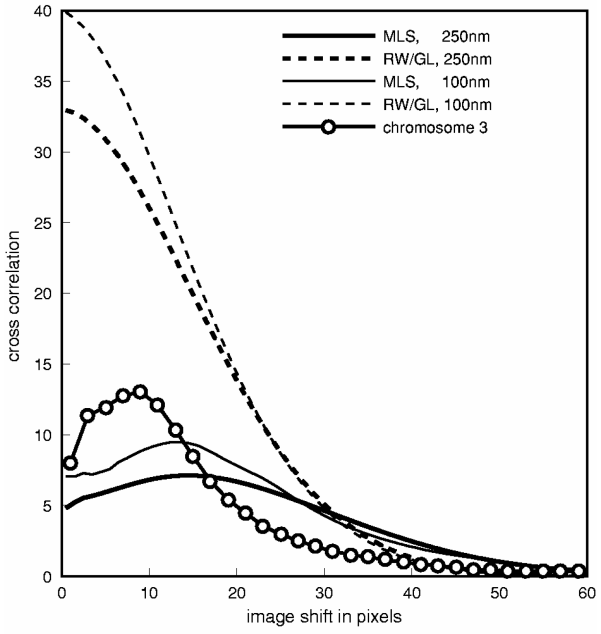

FIG. 7. (a) Overlap of chromosome-3 p and q arms predicted by the MLS model (solid lines) and the RW-GL model (dashed lines). The observed overlap (circles, from [40]) is most similar to the MLS-model data. The x-y scaling of the data depends on the ratio of chromosome to image size, which is different for both models and the experimental data. Thick lines indicate a simulated resolution of 250 nm and thin lines 100 nm.

Computed correlations for about 100 experimentally observed chromosome-3pq and -6pq configurations [40] resulted in about the same background levels. By increasing the threshold to $I_t = 10\%$ of the maximal intensity the background correlation signal was completely removed and the resulting correlation signal (Fig. 7, circles) is most similar to the MLS-model chromosomes. However, a perfect match of all the correlation functions cannot be expected because the scaling of both axes depends on the ratio of domain and image size, which was different for all three data sets.

Using the threshold derived from the experimental correlation function, we also computed the overlap O of both the observed and the simulated images, which is defined as

$$O = \frac{\sum_{\text{all pixels}} I_{\text{and}}}{\sum_{\text{all pixels}} I_{\text{or}}}, \tag{16}$$

i.e., by all pixels I_{and} with both color intensities above the threshold and I_{or} with any of the channels above the threshold:

$$I_{\text{and}} = \begin{cases} 1 & \text{for } I_1 > I_t, \quad I_2 > I_t \\ 0 & \text{otherwise,} \end{cases}$$

$$I_{\text{and}} = \begin{cases} 1 & \text{for } I_1 > I_t \text{ or } I_2 > I_t \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

For the MLS-model chromosome 3 and a resolution of 250 nm we find an overlap of about 6%, in agreement with the experimental overlap of 7% for chromosome-3 arm domains [40]. The RW-GL model results in a three times broader distribution with a mean overlap of 15%. In addition to contributions from the limited microscopical resolution some of the experimentally observed overlap can be caused by arm probes, which may partly cover the same region.

The notion of a highly compartmentalized chromosome territory structure predicted by the MLS model and verified by chromosome-3 arm observations is valid even on the level of subcompartments. Using a simulated 250-nm resolution for modeled chromosome-15 territories, the mean overlap of R and G subcompartments is about 7% for the MLS model and 20% for the RW-GL model. If the computation is restricted to the central core of the chromosome (about half the linear size), the mean overlap increases to about 10% and 34%, respectively. Hence the average overlap of band domains computed for the RW-GL model is about three times higher than for the MLS model, reflecting the entirely different internal structure of chromosome territories predicted by the two models, which was found already for the overlap of chromosome arms and the density profiles. The experimentally measured overlap of early and late replicating domains (i.e., R and G subcompartments) of chromosome-15 territories is about 4% [46].

Moreover, MLS-model estimates for subcompartment volumes agree with experimental data. For chromosome 15 in G1 about $4.2 \pm 1.0\ \mu\text{m}^3$ early and about $3 \pm 1\ \mu\text{m}^3$ late replicating chromatin per chromosome were measured [46]. Using the same image analysis software 27 simulated MLS-model chromosomes were analyzed [46]. In $G1$ the mean R-subcompartment volume was $3.0 \pm 0.7\ \mu\text{m}^3$ and the mean G-subcompartment volume was $2.5 \pm 0.8\ \mu\text{m}^3$. Additional uncertainties of about $1\ \mu\text{m}^3$ may be caused by approximations of the simulated vs experimental painting and imaging. The modified MLS model including some giant loops results in 2.5 times higher volumes of R subcompartments. For the RW-GL model volumes of both R and G bands are about 2.5 times larger than observed.

IV. CONCLUSION AND DISCUSSION

The agreement of the MLS-model predictions with experimental results demonstrates that simulations of polymer models are well suited to study the three-dimensional organization of the human genome in interphase. However, a comparison with many different experiments is required because measurements are indirect and the structures cannot be deduced uniquely. Mean interphase distances in the (10–200)-Mbp range, for example, can be reproduced by the MLS model, the RW-GL model, and a modified RW-GL model [47]. Additional measurements of distances below 5 Mbp and the overlap of chromosome arm and subcompartments using structure preserving preparation conditions exclude models with giant loops and without an excluded-volume interaction, i.e., strongly intermingled fibers.

The MLS model is based on very few parameters only. Therefore, it cannot represent the specific details of a certain chromosome in a certain cell cycle and differentiation state. Not all of the subcompartments may be in a decondensed or opened state; some may be condensed next to the connecting chromatin fragments between the subcompartments. This would lead to a much thicker fiber with a diameter of about 90–150 nm, similar to the findings of Belmont and Bruce

[4]. The opening of the subcompartments may also be a dynamic process related to transcription. Chromosome structure changes dramatically by the inhibition of RNA polymerase II [48] and regional differences in the compacting of chromatin were found [49]. Further experiments are required for such a refinement of the MLS model.

Our results support the hypothesis that chromosome territories have an organized internal structure. Further studies are required to determine the degree of this order, especially for smaller structures below 1 Mbp. The modeling and explanation of this order, as well as the dynamics of chromosome territory organization and distribution, must consider interactions between chromosome territories and with other nuclear organelles [50]. However, essential structural features of chromosome territory and band domain organization can be described without the need for an organization by additional mechanisms, e.g., a rigid nuclear matrix.

ACKNOWLEDGMENTS

This work was supported by BMBF Grant No. 01 KW 9620 (German Human Genome project). We thank M. Tewes for suggesting overlap computations by FFT and correlation functions, H. Bornfleth for computing subcompartment volumes with his image analysis software for a subset of our simulated chromosome 15 images, and G. Wedemann for many encouraging discussions.

[1] G. Felsenfeld, Cell **86**, 13 (1996).
[2] K. van Holde and J. Zlatanova, J. Biol. Chem. **270**, 8373 (1995): C. L. Woodcock and R. A. Horowitz, Trends Cell Biol. **5**, 272 (1995).
[3] R. van Driel *et al.*, Int. Rev. Cytol. **162A**, 151 (1995).
[4] A. S. Belmont and K. Bruce, J. Cell Biol. **127**, 287 (1994).
[5] L. Manuelidis, Science **250**, 1533 (1985).
[6] C. Cremer *et al.*, Hum. Genet. **80**, 235 (1988).
[7] P. Lichter *et al.*, Hum. Genet. **80**, 224 (1988).
[8] J. S. Heslop-Harrison and M. D. Bennet, Trends Genet. **6**, 401 (1990).
[9] D. E. Comings, Am. J. Hum. Genet. **20**, 440 (1968).
[10] P. Hahnfeldt *et al.*, Proc. Natl. Acad. Sci. USA **90**, 7854 (1993).
[11] J. Y. Ostashevsky and C. S. Lange, J. Biomol. Struct. Dyn. **11**, 813 (1994).
[12] G. van den Engh, R. Sachs, and B. J. Trask, Science **257**, 1410 (1992).
[13] R. K. Sachs, G. van den Engh, B. Trask, H. Yokota, and J. E. Hearst, Phys. Rev. A **92**, 2710 (1995).
[14] C. C. Robinett *et al.*, J. Cell Biology **135**, 1685 (1996).
[15] L. Manuelidis, Hum. Genet. **71**, 288 (1990).
[16] P. R. Cook, J. Cell. Sci. **108**, 2927 (1995).
[17] T. A. Okada and D. E. Comings, Chromosoma **72**, 1 (1979).
[18] T. Cremer *et al.*, *Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1993), Vol. LVIII.
[19] A. Kurz *et al.*, J. Cell Biol. **135**, 1195 (1996).
[20] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell*, 2nd ed. (Garland, New York, 1989).
[21] A. Wolffe, *Chromatin*, 2nd ed. (Academic, New York, 1995).
[22] B. Duplantier, G. Jannink, and J. L. Sikorav, Biophys. J. **69**, 1596 (1995).
[23] J. S. Heslop-Harrison *et al.*, Hum. Genet. **84**, 27 (1989).
[24] R. Eils *et al.*, J. Cell Biol. **135**, 1427 (1996).
[25] H. Yokota, G. van den Engh, J. E. Hearst, R. K. Sachs, and B. J. Trask, J. Cell Biol. **130**, 1239 (1995).
[26] Y. Saitoh and U. K. Laemmli, Cell **76**, 609 (1993).
[27] W. A. Bickmore and K. Oghene, Cell **84**, 95 (1996).
[28] K. J. Pienta and D. S. Coffey, J. Cell Sci. Suppl. **1**, 123 (1984).
[29] H. H. Q. Heng, J. W. Chamberlain, X.-M. Shi, B. Spyropoulos, L.-C. Tsui, and P. B. Moens, Proc. Natl. Acad. Sci. USA **93**, 2795 (1996).
[30] U. Francke Cytogenet. Cell Genet. **65**, 206 (1994).
[31] J. J. Yunis, Hum. Genet. **56**, 239 (1981).
[32] D. L. Ermak and J. A. McCammon, J. Chem. Phys. **69**, 1352 (1978).
[33] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
[34] N. Madras and A. D. Sokal, J. Stat. Phys. **50**, 109 (1988).
[35] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford Science, New York, 1987).
[36] S. Plimpton, J. Comput. Phys. **117**, 1 (1995); D. M. Beazley, P. S. Lomdahl, N. Gronbech-Jensen, R. Giles, and P. Tamayo, *Annual Reviews in Computational Physics* (World Scientific, Singapore, 1995), Vol. 3.
[37] C. Münkel and D. W. Heermann, Phys. Rev. Lett. **75**, 1666 (1995).
[38] P. Lacroute and M. Levoy, in *Computer Graphics Proceedings ACM SIGGRAPH 1994*, edited by A. Glassner (ACM Press, Orlando, 1994), p. 451.
[39] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1995).
[40] S. Dietzel *et al.*, Chromosome Research (to be published).
[41] J. B. Lawrence, R. H. Singer, and J. A. McNeil, Science **249**, 928 (1990).
[42] J. A. Warrington and U. Bentsson, Genomics **24**, 395 (1994).
[43] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[44] A. F. Dernburg, K. W. Broman, J. C. Fung, W. F. Marshall, J. Philips, D. A. Agard, and J. W. Sedat, Cell **85**, 745 (1996).
[45] A. Visser (private communication).
[46] D. Zink *et al.* (unpublished).
[47] B. Liu and R. K. Sachs, Bull. Math. Biophys. **59**, 325 (1997).
[48] T. Haaf and D. C. Ward, Exp. Cell Res. **224**, 163 (1996).
[49] H. Yokota, M. J. Singer, G. J. van den Engh, and B. J. Trask, Chromosome Res. **5**, 157 (1997).
[50] C. Münkel *et al.*, Bioimaging **3**, 108 (1995).