Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 2710–2714, March 1995
Biochemistry

A random-walk/giant-loop model for interphase chromosomes

(chromatin/fluorescent *in situ* hybridization/cell nucleus ultrastructure/polymers)

R. K. SACHS*, G. VAN DEN ENGH†, B. TRASK†, H. YOKOTA†, AND J. E. HEARST‡

Departments of *Mathematics and ‡Chemistry, University of California, Berkeley, CA 94720; and †Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195

Communicated by Alexandre J. Chorin, University of California, Berkeley, CA, November 21, 1994 (received for review October 17, 1994)

ABSTRACT **Fluorescence *in situ* hybridization data on distances between defined genomic sequences are used to construct a quantitative model for the overall geometric structure of a human chromosome. We suggest that the large-scale geometry during the G_0/G_1 part of the cell cycle may consist of flexible chromatin loops, averaging ≈3 million bp, with a random-walk backbone. A fully explicit, three-parametric polymer model of this random-walk/giant-loop structure can account well for the data. More general models consistent with the data are briefly discussed.**

A human chromosome is a very large molecule. Its DNA strand has in the order of 100 million base pairs (Mbp) arrayed along its contour and has a M_r of $\approx 10^{11}$ Da. Quantitative information on mammalian chromosome geometry during the interphase part of the cell cycle is very extensive for scales <0.01 Mbp (1, 2) but not for larger scales. At the level of ≈0.001–0.01 Mbp the DNA is associated with proteins to form a chromatin fiber ≈30 nm in diameter (1, 2); at scales of ≈0.1 Mbp the chromatin may form loops (2). Very little is known numerically about the larger-scale geometry, comprising >3 orders of magnitude (0.1–300 Mbp), where the difficulty of following the chromatin fiber as it winds and twists its way within the interphase cell nucleus has crippled quantitative analyses. Yet, large-scale geometric structure of chromosomes influences essential cellular processes such as DNA replication and transcription (2), as well as many specialized functions such as repair or misrepair of ionizing-radiation-produced DNA damage (3, 4).

Recently, van den Engh *et al.* (5) used fluorescence *in situ* hybridization data to quantify some intermediate-scale properties of interphase chromosomes. These investigators measured physical distances between pairs of fluorescently marked specific DNA sequences on human chromosome 4 in fibroblast cells fixed on microscope slides. The observations were made for cells in the G_0/G_1 phase of the cell cycle, the period between mitosis and the onset of DNA replication. Probe pairs having genomic separations from ≈0.1 Mbp to ≈4 Mbp were analyzed. A major conclusion was that, on scales from 0.1 Mbp to 1.5 Mbp, chromatin geometry corresponds to a simple random walk. Observed deviations from random-walk behavior at larger genomic separations could be explained by a polymer model in which the DNA of any one chromosome is confined to a spherical subvolume of the interphase nucleus (6). An alternative suggestion was that the deviations were due to "giant" loops, several Mbp in length (7)—i.e., far bigger than the more familiar ≈0.1-Mbp loops whose properties are summarized, e.g., by Tsanev *et al.* (2).

H.Y. *et al.* (unpublished data; H.Y., G.v.d.E., and B.T.) have now extended such fluorescent probe-pair data to much larger scales, up to the full length of a chromosome. Physical distances were obtained for >100 probe pairs, having genomic separations from 0.15 Mbp to 190 Mbp and located on human chromosomes 4, 5, or 19. These new data can be used to test the models of large-scale chromatin geometry. Our analysis will presuppose that the distributions observed in the fixed samples of Yokota *et al.* (8) reflect the distribution of marker pairs in the living cell. On this assumption the data show there are two, and probably only two, levels of chromatin structure over the range 0.15–200 Mbp, with the chromatin at genomic separations greater than a few Mbp not spreading out as fast as one would expect from its behavior in the 0.15- to 1.5-Mbp range. We shall argue that the simplest interpretation of the data is to assume that there are flexible chromatin loops averaging ≈3 Mbp, with their base points along a random walk. The heart of the argument is given by Eqs. **10–12** in the *Results* section.

DATA

Figs. 1 and 2 compare results of H.Y. *et al.* (unpublished data) with a model to be introduced later. Relationships between mean-square physical distance and genomic separation fall into two markedly different regimes, with a sharp transition at several Mbp (Fig. 1). For values of genomic separation <1.5 Mbp, the points lie approximately on a straight line (Fig. 1*A*), corresponding to random-walk behavior for chromatin, in agreement with earlier data (5). Genomic sites separated by 10–190 Mbp also show an approximately linear relation but with a much smaller slope (Fig. 1*B*). Moreover, for these large genomic separations, the statistical distribution of distances for a given probe pair also corresponds to random-walk behavior (Fig. 2). We now present a mathematical model, used to interpret Figs. 1 and 2 as well as other results of H.Y. *et al.* (unpublished data).

MATHEMATICAL METHODS

Polymer Models. As in other recent treatments (5–7) we consider polymer models for the chromatin at scales ≥0.1 Mbp. The idea behind polymer models is that by ignoring small-scale intricacies one can arrive at comparatively very simple overall statements about average large-scale behavior in situations where random influences smooth out details (9, 10).

As shown in Fig. 3, the models use a sequence of "beads" located at equal genomic separations along the chromatin (10). The beads are primarily a bookkeeping device, and their postulated genomic separation can be selected to fit the situation of interest. For our purposes, the separation should not be much larger than the resolution of the data, ≈0.1 Mbp, but should be large compared with the statistical segment of the worm-like chromatin coil that connects consecutive beads. Free DNA has a statistical segment of $\approx 10^{-3}$ Mbp (11), and the 30-nm fiber appears to have a statistical segment in the 0.01- to 0.1-Mbp range (6, 7). Therefore we assign a separation

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: Mbp, million base pairs.

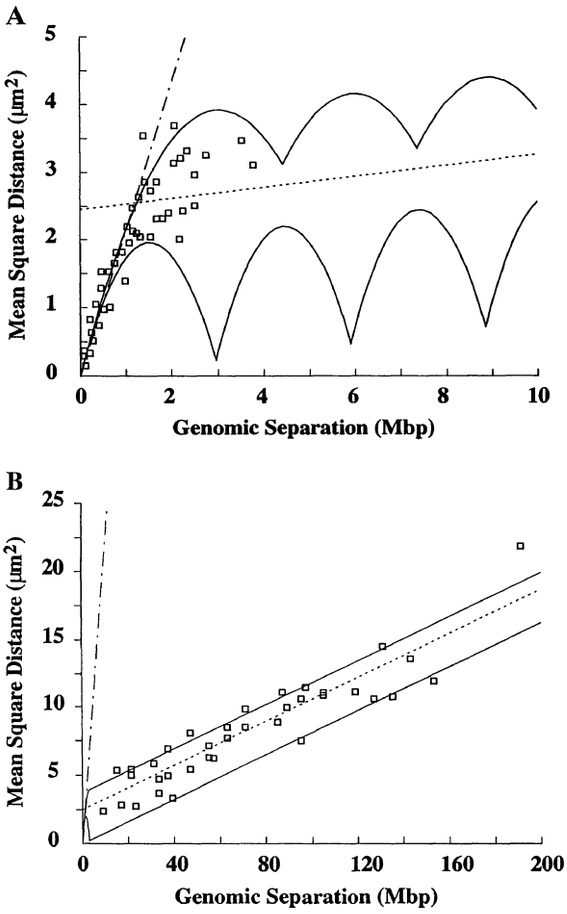

FIG. 1. Physical distances and genomic separations. Graphs show mean-square two-dimensional distance $\langle r^2 \rangle$ between two DNA probes as a function of their genomic separation along human chromosome 4. All data points are from H.Y. *et al.* (unpublished data), "preparation 384." Nineteen different DNA sequences were measured in 80 different pairs, 44 for genomic separations of 0.15–3.5 Mbp (*A*) and 36 for separations of 10–190 Mbp (*B*). For each data point ≈150 replicate measurements were made (unpublished data). SEMs (not shown) are ≈4%. In *A* and *B* the broken (– ·) line shows the initial slope in *A*, and the dotted line is the best linear fit to the data points in *B*; the difference in slope is ≈20-fold. In *A*, the solid, curved lines are upper and lower bounds predicted by the random-walk/giant-loop model (see text). The solid lines in *B* are envelopes of the solid lines shown in *A*.

of 0.1 Mbp to the genomic separation between consecutive beads; by rescaling (10) one could use other, similar values without substantive changes in any of the arguments or results below. Structure on genomic separations significantly >0.1 Mbp, such as structure responsible for the transition at several Mbp in Fig. 1, can be modeled by interactions between beads that are not consecutive (see *Results*).

We follow the notation of Doi and Edwards (10). Cartesian coordinates of the beads are denoted by $\mathbf{X}_i$, i = 0, . . ., N (Fig. 3). The Z direction is taken perpendicular to the microscope slide. The data concern only the (X, Y) plane. Cartesian components and projections in the (X, Y) plane are written as follows:

$$\mathbf{X}_j = (X_j, Y_j, Z_j) = (\mathbf{x}_j, Z_j), \qquad \text{i.e., } (X_j, Y_j) \equiv \mathbf{x}_j. \qquad [1]$$

Equilibrium statistical properties of polymers are often represented as properties of a mechanical system with negligible kinetic energy (10). The probability density $G(\mathbf{X}_0, \ldots, \mathbf{X}_N)$ for having a bead configuration $\mathbf{X}_0, \ldots, \mathbf{X}_N$ can then be regarded as a Boltzmann distribution (10):

$$G(\mathbf{X}_0, \ldots, \mathbf{X}_N) = C \exp(-U/k_BT). \qquad [2]$$

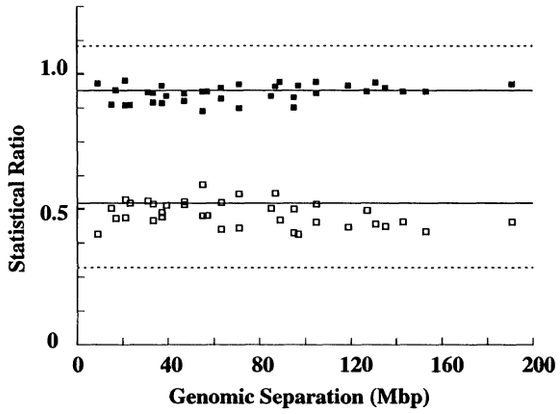

FIG. 2. Statistical ratios. For the data points in Fig. 1*B*, open squares show the ratio of *sd* in distance r to the mean, and the closed squares show the ratio of median to mean. Solid lines show the ratios predicted by the flexible-backbone/giant-loop model and do not depend on any adjustable parameters. Dotted lines, shown for comparison, are the ratios for a randomly oriented, rigid rod.

Here C is the normalization constant, $U(\mathbf{X}_0, \ldots, \mathbf{X}_N)$ is the potential energy of the system of beads, k_B is Boltzmann's constant, and T is the temperature.

Distance Probability Densities for a Fluorescent Probe Pair. A given pair of fluorescent probes can be taken to correspond to a particular pair of beads, numbered i and j, respectively. The observations concern the two-dimensional projected scalar physical distance r between the probes, i.e.,

$$r = [(X_i - X_j)^2 + (Y_i - Y_j)^2]^{1/2} = \|\mathbf{x}_i - \mathbf{x}_j\|. \qquad [3]$$

To analyze the data, one needs the distance probability density $P(r)$ for r. $P(r)$ can be obtained by first finding the joint probability density $P_2(\mathbf{x}_i - \mathbf{x}_j) \equiv P_2(u, v)$ for the two Cartesian intervals u and v defined by $u = X_i - X_j$ and $v = Y_i - Y_j$. P_2 in turn can be obtained by integrating G in Eq. **2** over all variables except u and v, i.e.,

$$P_2(\mathbf{x}_i - \mathbf{x}_j) = C \int ds\, dw\, dZ_i\, dZ_j \int d\mathbf{X} \exp(-U/k_BT), \qquad [4]$$

where $s = \frac{1}{2}(X_i + X_j)$, $w = \frac{1}{2}(Y_i + Y_j)$, and $d\mathbf{X}$ denotes integration over $3 \times (N - 1)$ Cartesian variables. The distance probability density $P(r)$ is then obtained by the Jacobian relation appropriate for two dimensions, i.e.,

$$P(r)\, dr = r\, dr \int_0^{2\pi} d\theta P_2(r\cos\theta, r\sin\theta). \qquad [5]$$

RESULTS

We now model chromosome geometry, first on intermediate scales of 0.1–1.5 Mbp and then on larger scales.

Intermediate Scales. The model used by van den Engh *et al.* (5) to analyze chromatin behavior at the 0.1- to 1.5-Mbp level is equivalent to taking $U = U_0$ in the probability density of Eq. **2**, where

$$U_0 = (\kappa/2) \sum_{j=1}^{N} \|\mathbf{X}_j - \mathbf{X}_{j-1}\|^2, \qquad [6]$$

with κ a nominal spring constant that elegantly summarizes all relevant smaller-scale chromatin structure (10). If $U = U_0$, the mean-square physical distance depends linearly on genomic separation (5, 10). The recent data of H.Y. *et al.* (unpublished data) confirm the earlier results that on scales of 0.1–1.5 Mbp this model is appropriate, in that points in the lower left of Fig.

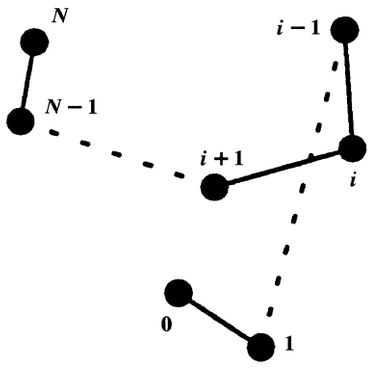

FIG. 3. Polymer models, showing the numbering of "beads," (solid circles), used to track the large-scale geometry of a chromosome. Beads 0 and N are at the chromosome ends (telomeres). $N \approx 1900$ for human chromosome 4. Chromatin connections between consecutively numbered beads are shown schematically as straight lines but are actually tortuous paths with a contour length very much larger than the bead-to-bead distance. With a genomic separation of 0.1 Mbp between consecutive beads, their rms distance, determined by Eqs. **11** and **12**, is $\approx 1/2$ μm. The intricate DNA–protein structure of the 30-nm chromatin fiber is too small-scale to appear (and appears in the mathematics of the text only via a single constant κ). Higher-order chromatin structure is modeled by additional interactions (not shown)—e.g., interactions between nonconsecutive beads (Fig. 4) and/or interactions of the beads with an extrachromosomal structure.

1*A* lie approximately on a straight line (whose slope is ≈ 2 μm^2/Mbp). Therefore we later take $U = U_0 + \hat{U}$ in Eq. **2**, where $\hat{U}$ is almost negligible at scales of <1 Mbp but is adjusted to model higher-order structure.

In the special case $U = U_0$, the probability density (2) is multivariate normal (i.e., Gaussian), and Eq. **4** shows that the two-dimensional probability density P_2 also has a Gaussian form—namely,

$$P_2(\mathbf{x}_i - \mathbf{x}_j) = [\pi\Gamma]^{-1}\exp[-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\Gamma], \quad \Gamma = \frac{k_B T|i-j|}{\kappa}. \quad [7]$$

The further integration specified by Eq. **5**, applied to the probability density of Eq. **7**, gives for the distance probability density $P(r)$ the form used by van den Engh *et al.* (5) in analyzing the data for length scales of 0.1–1.5 Mbp—namely,

$$P(r) = (r/\Gamma)\exp[-r^2/2\Gamma]. \quad [8]$$

Rayleigh Probability Densities. Any probability density of the form of Eq. **8**, with Γ independent of r, is referred to as a Rayleigh probability density (12). A Rayleigh probability density leads to standard relations for means, mean squares, medians, etc. For example, denoting averages by $\langle \ldots \rangle$, one has the following results for the mean $\langle r \rangle$ and for the mean square $\langle r^2 \rangle$:

$$\langle r \rangle = \int_0^\infty rP(r)dr = [\pi\Gamma/2]^{1/2}; \quad \langle r^2 \rangle = \int_0^\infty r^2P(r)dr = 2\Gamma. \quad [9]$$

For the standard deviation sd this gives $sd/\langle r \rangle = [(4/\pi) - 1]^{1/2} \approx 0.52$. Similarly, one can derive median/$\langle r \rangle \approx 0.94$. The solid lines in Fig. 2 show these values.

For comparison with the Rayleigh probability density, we consider the probability density for r when one has two probes fixed on a rigid, randomly oriented rod. Such a model could approximate a chromosome with a rigid backbone in a spherical nucleus, so the comparison gives some insight into whether the results in Fig. 2 are informative. The probability that a rigid rod makes angle θ with the Z axis is $\sin\theta$, and from this probability the statistical ratios shown as dotted lines in Fig. 2 can be worked out.

Large Scales. We now consider chromosome geometry for scales of 2–200 Mbp. One striking feature of the experimental observations for large genomic separations is that various statistical ratios involving r approximate the Rayleigh values, independent of genomic separation (H.Y. *et al.*, unpublished data); some of these data are shown in Fig. 2. However, assuming Rayleigh probability densities with the additional property of Γ being proportional to genomic separations $|i - j|$ as in Eq. **7** would imply, via Eq. **9**, a linear dependence of $\langle r^2 \rangle$ on genomic separation over the entire range, whereas Fig. 1 shows biphasic behavior. We reasoned that by introducing additional interactions between distant beads, but continuing to assume U quadratic in the Cartesian coordinates, one could obtain Rayleigh probability densities for each probe pair while allowing more general dependence of $\langle r^2 \rangle$ on genomic separation.

The sharp transition of the curve in Fig. 1 at several Mbp suggests that there are at least two levels of chromosome structure on scales of 0.1 Mbp or more. Significantly, the apparent absence of additional structure in the curve for the range 10–200 Mbp also suggests that perhaps there are *only* two levels. The change in slope near several Mbp in Fig. 1 corresponds to a change from a loose random walk at smaller scales to a tighter random walk at the largest scales. This transition from looser to tighter suggests for a chromosome some bending, or looping (7), or spiraling (13), or volume constraint (6), or statistical clumping structure (14), with a scale of a few Mbp.

Giant-loop models (7) are the simplest way to model such distance dependence. The loops could be imposed by protein links between chromosome locations separated by millions of base pairs. For concreteness we shall first present a very specific model (Fig. 4), and then outline a few of the (many) alternatives or generalizations that differ in comparatively minor ways.

A Random-Walk/Giant-Loop Model. The model to be presented here has three adjustable parameters, corresponding to the slopes of the two straight lines in Fig. 1*A* and the y intercept of the dotted curve.

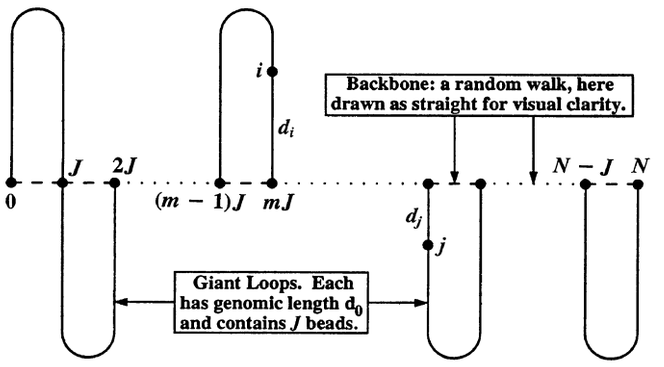

FIG. 4. A flexible-backbone/giant-loop model. The figure indicates the mathematical conventions used, but not the actual geometry, which has far more randomness and flexibility. As an idealization, all loops are taken to have equal genomic length. For $J \approx 30$, the beads numbered 0, J, $2J$, ..., $(m - 1)$ J, mJ, ..., $N - J$, N are loop-attachment points. Connections between consecutive loop-attachment points are shown schematically as dotted lines. A model with short (<0.2 Mbp) stretches of "linker chromatin" between loops, analogous (on a much larger scale) to linker DNA between nucleosomes, would give very similar results. The schematic shows five giant loops out of a total of N/J (≈ 65 for human chromosome 4). Beads numbered i and j are fluorescently labeled. The model predicts the statistical distribution of physical distances between such fluorescent probe pairs. d_i and d_j are the genomic separations of the fluorescent probes from the nearest loop-attachment points lying between the ith and jth beads.

For a formal definition of the model, suppose that U in Eq. **2** contains, in addition to the terms involving consecutive beads, specified by the potential U_0 of Eq. **6**, harmonic interaction terms between bead pairs separated by large genomic intervals (Fig. 4). Specifically, suppose that for some integer J, N/J is an integer, the total number of giant loops. Suppose the end bead numbered 0 interacts with the Jth bead, the Jth bead interacts with the $2J$th bead, and so on up to the interaction of the $N - J$th bead with the bead, numbered N, at the other end (Fig. 4). Thus the beads numbered $0, J, 2J, \ldots, N$ are loop-attachment points, and a "backbone" is determined by flexible connections between consecutive loop-attachment points (Fig. 4). If the interaction is the same for each pair of consecutive loop-attachment points, U is given by

$$U = U_0 + (\hat{\kappa}/2) \sum_{m=1}^{N/J} \|\mathbf{X}_{mJ} - \mathbf{X}_{(m-1)J}\|^2, \qquad [10]$$

where U_0 is the nearest-neighbor term of Eq. **6** and $\hat{\kappa} \geq 0$ is an effective spring constant for the backbone. Eqs. **2**, **6**, and **10** define the random-walk/giant-loop model mathematically.

For comparison with the data in Fig. 2, note that if Eq. **10** holds, the probability density G defined in Eq. **2** is a multivariate normal function of the $3N + 3$ Cartesian coordinates $X_0, Y_0, \ldots, Y_N, Z_N$. By standard theorems on multivariate normal distributions (15), P_2 ($\mathbf{x}_i - \mathbf{x}_j$) is a (nondegenerate) multivariate normal function of the two variables u and v defined above Eq. **4**. Eq. **10** also implies isotropy—i.e., no preferred directions. Isotropy implies that u and v are uncorrelated, have zero means, and have equal variances. Therefore, for some constant Γ, the Gaussian form Eq. **7** must hold. Eq. **7** directly implies the Rayleigh form, Eq. **8**. Thus the random-walk/giant-loop model, Eq. **10**, is consistent with the data shown in Fig. 2 and other data (H.Y. *et al.*, unpublished data) that suggests Rayleigh distance distributions.

To compare with the data of Fig. 1, one only needs to determine the dependence of Γ (and thus of $\langle r^2 \rangle$) on the genomic locations of the ith and jth beads, using the probability density given by Eqs. **2**, **6**, and **10**. Here i and j need not be loop-attachment points—i.e., need not be integer multiples of J (Fig. 4). After evaluating the relevant integrals by a long calculation, omitted here, it turns out that the results are identical to results obtained by the following, rather amusing, trick. One can find an "equivalent spring constant," call it κ_{ij}, for the "equivalent spring" between the ith bead and the jth bead, using the usual elementary rules for combining springs in parallel (add the constants) or in series (use the reciprocal of the sum of reciprocals). It is then found from Eqs. **4**, **5**, and **9** that $\langle r^2 \rangle$ is determined simply by the usual equipartition rule for two degrees of freedom—i.e., $\frac{1}{2}\, \kappa_{ij} \langle r^2 \rangle = 2 \times (\frac{1}{2}\, k_B T)$, implying $\langle r^2 \rangle = 2k_B T/\kappa_{ij}$.

If the ith bead and jth bead are not on the same giant loop—i.e., if there is at least one integer of the form mJ between i and j (Fig. 4), integration (or combining springs) gives

$$\langle r^2 \rangle = \hat{S}d + S[d_i(d_0 - d_i) + d_j(d_0 - d_j)]. \qquad [11]$$

Here d is the genomic separation between the probes, d_0 is the genomic length of a giant loop, and d_i and d_j are intraloop genomic locations. If the ith and jth bead are on the same giant loop, Eq. **11** still applies, provided we set $d_i = d$ and $d_j = 0$. In the equation, S and $\hat{S}$ are scale factors, having dimensions of μm^2. Specifically, $\hat{S} = k_B T/(A + \hat{\kappa} d_0)$ and $S = \hat{\kappa}\hat{S}/A$, where $A \equiv d_0 \lambda / J$. The three essential parameters of the model are S, $\hat{S}$, and a giant loop genomic size d_0; these are adjusted from the data in the way explained above Eq. **12**.

Eq. **11** shows that the mean-square physical distance depends on the intraloop locations d_i and d_j, not just on the separation d. In practice d_i and d_j are not known, so in Fig. 1*A* two solid curves are drawn to show extremes of the theoretical values. The lower solid curve gives the minimum in Eq. **11** for a given d, which occurs when $d_i = 0$ (or $d_j = 0$). The upper solid curve in Fig. 1*A* gives the maximum, which occurs if $d_i = d_j \leq 3d_0/4$. For a given d the other theoretical possibilities lie between these two extremes. The upper and lower solid straight lines in Fig. 1*B* are corresponding envelopes, determined respectively by the points where $d_i = d_j = \frac{1}{2}\, d_0$ or $d_i = d_j = 0$. The detailed theoretical bounds for Fig. 1*B* are actually curves that continue the solid curves of Fig. 1*A*, so data points lying between the solid straight lines in Fig. 1*B* may be outside the detailed theoretical bounds. However, the latter are extremely sensitive to changes in the model's adjustable parameters, uncertainties in the empirical values of d, or possible variability in giant-loop size, so the more robust envelope solid straight lines in Fig. 1*B* are perhaps more informative as bounds.

To estimate the three adjustable parameters empirically we regard the dotted curve in Fig. 1*B* as an average over internal loop locations, taking d_i and d_j independently and uniformly distributed over the interval $[0, d_0]$. Averaging gives $\langle r^2 \rangle = \hat{S}d + \frac{1}{3}\, S d_0^2$—i.e., the parameters $\hat{S}$ and $\frac{1}{3}\, S d_0^2$ are the slope and the y intercept, respectively, of the dotted curve in Fig. 1*B*. A third numerical relation is obtained by finding the best fit of Eq. **11** to the initial slope (Fig. 1*A*, broken line). Evaluating the parameters in this way gives

$$S \approx 0.83\ \mu m^2/Mbp^2, \quad \hat{S} \approx 0.081\ \mu m^2/Mbp,$$
$$d_0 \approx 3\ Mbp. \qquad [12]$$

With the parameter values given in Eq. **12**, the model, Eq. **10**, reproduces the main features of the data in Fig. 1, including the steep initial linear portion, the transition at several Mbp, and the more gradual linear increase at large separations. The fact that the data-point cloud for 1.5–3.5 Mbp (Fig. 1*A*) lies generally within the predicted range (solid curves) is support for the model because the giant-loop size, and thus the curvature of the solid curves, was determined from the average behavior of the large-scale data (dotted curve) and the initial slope of the smaller-scale data (broken curve), without use of data points in the region of 1.5–3.5 Mbp. The fact that the data points in Fig. 1*B* are generally within the solid curves also lends support to the model, as none of the parameter adjustment used the observed width of the cloud.

Other Models. Most of the alternatives to the random-walk/giant-loop model are more complicated and/or involve additional adjustable parameters. We take the view here that models involving additional adjustable parameters are too detailed for the present data set. For example, it would be more realistic to assume that the loops are not all exactly the same size, but a model taking variations into account would involve at least one more adjustable parameter—e.g., the variance in loop size.

A rather general model can be given, which allows for many other kinds of large-scale chromosome structure and is consistent with the statistical ratio data shown in Fig. 2. Mathematically, the relevant assumption is that, for a suitably chosen origin of Cartesian coordinates, the potential energy U in Eq. **2** has the specific form

$$U = \sum_{i=0, j \neq i}^{N} A_{ij} \|X_i - X_j\|^2 + \sum_{i=0}^{N} A_i \|X_i\|^2, \qquad [13]$$

with the A_{ij} and the A_i nonnegative constants and the matrix A_{ij} irreducible in the Perron–Frobenius sense (16). The random-walk/giant-loop model, Eq. **10**, is a special case of Eq. **13**. By a routine extension of the arguments given below Eq. **10**, Eq. **13** is seen to imply Rayleigh distributions for each probe

pair, as suggested by the data in Fig. 2. By selecting suitable coefficients A_{ij}, Eq. **13** can be used to model chromosomes having linker chromatin stretches between successive giant loops, analogous (on a much larger scale) to linker chromatin between nucleosoms, or to model chromosomes having multiple intrastrand connections. Moreover, if $A_i \neq 0$ for some i, Eq. **13** describes a chromosome whose higher-order structure is modified by flexible tethers to an extrachromosomal organizing center small compared with the chromosome. However, significant tethering to a rigid extrachromosomal scaffold of micrometer dimensions is not compatible with Eq. **13** and for large genomic separations gives pairwise distance probability distributions similar to those for a randomly oriented rigid rod, in contrast to the data (Fig. 2).

In deriving Rayleigh distributions, we have implicitly assumed that the probability density (Eq. **2**) is the same in all cells and at all sampled times during G_0/G_1, corresponding to stable, DNA-sequence-specific higher-order structure. For example, if large-scale structural features were different from cell to cell and Eq. **13** held for the interactions in each cell separately, there would be extra variance, giving a larger ratio of sd/mean than shown by the solid curve in Fig. 2.

DISCUSSION

This paper considered data on distances between defined pairs of points on a chromosome during cell-cycle interphase. The data are relevant to the geometric structure of mammalian chromatin on a scale larger than has hitherto been analyzed quantitatively, corresponding to genomic separations from 0.15 to 190 million bp. A three-parametric polymer model for flexible loops on a flexible backbone fits the data well. The loops are "giant," ≈3 Mbp, where the numerical value is comparatively sensitive to details of the fitting procedure. Alternatives and generalizations, for example models with multiple backbones or with flexible tethers to one small extrachromosomal "organizing center," were briefly discussed. The data speak against systematic attachments to a large stiff scaffold.

A stringent test of the presence of giant loops with DNA-sequence-specific attachments is to search for regions of "doubling back." For example, suppose one probe happens to be located near a loop-attachment point. Then for a series of probes that are slightly less than one giant loop away from the fixed probe increasing genomic separation corresponds to decreasing mean-square physical distance (Fig. 1*A*). H.Y. *et al.* (unpublished data) present direct evidence for doubling back in one case.

On the random-walk/giant-loop model, the loops and the backbone are essentially random walks extending over micrometer dimensions. The mean-square end-to-end distance for chromosome 4 is given by Fig. 1*B* as ≈20 μm^2. The mean-square "diameter" of a giant loop—i.e., the value obtained by setting $d_i = d = \frac{1}{2} d_0$, $d_j = 0$ in Eq. **11**—is $\frac{1}{2} \hat{S} d_0 + \frac{1}{4} S d_0^2 \approx 0.92$ μm^2. Chromosome size, as judged for example by the radius of gyration, is very significantly smaller than the cell nucleus size (H.Y. *et al.*, unpublished data). This picture is consistent with chromosomes having "territories" within the cell nucleus, as suggested by whole-chromosome painting (17) or indirectly by radiobiological evidence (3, 4, 8).

In summary, our results show that the methods of polymer theory, based on looking for simple average properties of an immensely complicated configuration, are an appropriate way to quantify large-scale chromosome geometry during interphase. The data suggest there are two and only two levels of chromosome structure at scales from 0.15 to 190 Mbp. The simplest model consistent with the data has, for the higher-order chromatin structure, flexible loops of several Mbp whose base points lie on a random walk.

We are grateful to E. Cheung for useful discussions. This study was supported in part by National Science Foundation Grant DMS-930-2704 (R.K.S.), National Institutes of Health Grant HG00256 (B.T.), Department of Energy Grant FG06-93ER61553 (B.T.), and National Institutes of Health Grant R01-GM47945-03 (J.E.H.).

1. Van Holde, K. E. (1989) *Chromatin* (Springer, New York).
2. Tsanev, R., Russev, G., Pashev, I. & Zlatanova, J. (1993) *Replication and Transcription of Chromatin* (CRC, Boca Raton, FL).
3. Savage, J. R. (1993) *Environ. Mol. Mutagen.* **22,** 234–244.
4. Brenner, D. J. & Sachs, R. K. (1994) *Radiat. Res.* **140,** 134–142.
5. van den Engh, G., Sachs, R. & Trask, B. (1992) *Science* **257,** 1410–1412.
6. Hahnfeldt, P., Hearst, J. E., Brenner, D. E., Sachs, R. K. & Hlatky, L. H. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 7854–7858.
7. Ostashevsky, J. Y. & Lange, C. S. (1994) *J. Biomol. Struct. Dyn.* **11,** 813–820.
8. Hlatky, L. R., Sachs, R. K. & Hahnfeldt, P. (1992) *Radiat. Res.* **129,** 304–308.
9. Cantor, C. R. & Schimmel, P. R. (1980) *Biophysical Chemistry* (Freeman, New York), pp. 979–1037.
10. Doi, M. & Edwards, S. F. (1988) *The Theory of Polymer Dynamics* (Oxford Univ. Press, Oxford).
11. Hearst, J. E. & Stockmayer, W. H. (1962) *J. Chem. Phys.* **37,** 1425–1433.
12. Trived, K. S. (1982) *Probability and Statistics, with Reliability, Queuing, and Computer Science Applications* (Prentice–Hall, Englewood Cliffs, NJ).
13. de la Tour, E. B. & Laemmli, U. K. (1988) *Cell* **55,** 937–944.
14. des Gennes, P. (1979) *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, NY).
15. Grimmett, G. R. & Stirzaker, D. R. (1992) *Probability and Random Processes* (Oxford Univ. Press, New York), 2nd Ed.
16. Fiedler, M. (1986) *Special Matrices and Their Applications in Numerical Mathematics* (Dordrecht, Boston).
17. Lichter, P., Cremer, C., Borden, J., Manuelidis, L. & Ward, D. C. (1988) *Hum. Genet.* **80,** 224–2348.