Dynamical Modeling of 3D Genome Organization in Interphase Budding Yeast

Naoko Tokuda
Department of Computational Science and Engineering,
Nagoya University, Nagoya 464-8603, Japan

Tomoki P. Terada
Department of Computational Science and Engineering,
Nagoya University, Nagoya 464-8603, Japan

Masaki Sasai[1]
Department of Computational Science and Engineering,
Nagoya University, Nagoya 464-8603, Japan,
School of Computational Sciences,
Korea Institute for Advanced Study, Seoul 130-012, Korea,
Okazaki Institute for Integrative Bioscience,
Okazaki 444-8787, Japan

[1]Corresponding author. Address: Department of Computational Science and Engineering, Nagoya University, Furo-cho Chikusa-ku, Nagoya 464-8603, Japan, Tel.: 81(52)789-4763, Fax: 81(52)789-3719

Abstract

Eukaryotic genome is organized in a set of chromosomes each of which consists of a chain of DNA and associated proteins. Processes involving DNA such as transcription, duplication and repair, therefore, should be intrinsically related to the three-dimensional (3D) organization of the genome. In the present paper, we develop a computational model of the 3D organization of the haploid genome of interphase budding yeast by regarding chromosomes as chains moving under the constraints of nuclear structure and chromatin-chromatin interactions. The simulated genome structure largely fluctuates with the diffusive movement of chromosomes. This fluctuation, however, is not completely random but parts of chromosomes distribute in characteristic ways to form 'territories' in nucleus. By suitably taking account of constraints arising from the data of the chromosome-conformation-capture measurement, the model explains the observed fluorescence data of chromosome distributions and motions.

Key words: Chromosome; Chromatin; Langevin dynamics; The Go-like potential; The 3C-based method

Introduction

Genome is not an abstract linear sequence but has a physical structure organized in the three-dimensional (3D) space. In eukaryotic cells, DNA folds hierarchically into several layers from chromatin to chromosome and to the whole genome, so that the DNA-related processes such as transcription, duplication, and repair should be affected or regulated by the 3D organization of the genome (1–4): It has been observed, for instance, that the disparate DNA elements co-localize in interphase nuclei to form 'transcription factories' (5). Such structure of interphase chromosomes, however, is not a frozen static configuration but is subject to the intense dynamical fluctuation.

In one view, this dynamical motion has been ascribed to the random movement of chromosome chains. In the human genome, for example, the observation based on the chromosome conformation capture (3C) method (6) has shown that the genome is organized as a fractal globule (7) whose features are well reproduced by the Monte Carlo simulation of the randomly moving non-specific polymer chains (6, 8). Rosa et al. (9, 10) have extended the worm-like chain (WLC) model of semiflexible polymers to describe chromosomes and shown that a non-specific kinetic effect of the Brownian motion of chromosome chains can separate different parts of the genome into 'territories'. Also in the genome of budding yeast, observation of the GFP-tagged loci has shown that each chromosome moves about 0.5μm during 10 seconds, almost a half of the radial length of the nucleus (11–13). The recent fluorescence measurement has shown that the positional distribution of telomeres widely spreads in the interphase yeast nucleus but is largely determined by the arm length of each chromosome (14, 15). These results suggest that the important features of interphase nuclei can be captured by regarding nuclei as solutions of non-specific polymers (9, 16).

In the other view, the genome structure has been regarded as an ensemble of configurations which are constrained by specific interactions between chromosomes and the nuclear structure (3, 17, 18) and also by specific interactions among chromosomes (1). In yeast, for example, centromeres are anchored to the spindle pole body (SPB), a protein structure embedded in the nuclear envelope (17, 19, 20). The repetitive elements of ribosomal DNA (rDNA) are confined in the nucleolus which is positioned at the opposite side of the nucleus from SPB (17, 20, 21). These constraints should break symmetry of the chromosome distribution to give rise to the 'Rabl-like' structure, which has been recently confirmed by the 3C-based measurement (22). By using the 3C-based methods, the frequency that two positions of chromosomes approach in proximity has been measured with the kilobase

pair (kbp) resolutions, from which 3D models of genome structure have been constructed for budding yeast (22) and fission yeast (23). It has been reported that for fission yeast thus observed proximity and the expression pattern of genes are correlated to each other, and those significantly associating genes frequently contain the same DNA motifs at their promoter regions (23). Tanizawa et al. have suggested that putative specific factors binding to these motifs are involved in defining the associations among genes (23). Upon induction of double-strand break (DSB), it has been observed that two loci co-localize to form a focus with proteins of the repair machinery (24). These examples have shown the importance of specific interactions between loci and between locus and a nuclear landmark (4).

By unifying these two views, the view based on the polymer dynamics and the view based on the specific interactions, it should become possible to examine intriguing and important questions on how the Rabl-like global structure is maintained or changed and how the specific local structures are formed or dissolved under the intense dynamical fluctuation. Computer modeling should be an efficient approach to explore this problem, and in this paper we introduce a computational model of the 3D genome organization in the interphase nucleus of haploid budding yeast. As has been pointed out by Grosberg et al. (7) and Marti-Renom and Mirny (25), this problem resembles in a sense the problem of protein folding: In these problems, both the polymer dynamics and the specific interactions are important, and hence we here borrow the idea of structural modeling from protein folding study (26).

The 3C-based method has provided information of the frequency that two sites in the genome come close to each other. From such frequencies, Duan et al. estimated mean distances between sites in the interphase haploid budding yeast (22). We use these distances to define the Go-like potential for genome. In the problem of protein folding, a simulated protein folds into the unique structure which minimizes the Go-like potential when the balance between temperature and the interaction energy in the model is favourable for folding (27, 28). In the condition which favors the more loosened structure, i.e., in high temperature or with the small interaction parameters, the protein chain does not settle in a folded structure, but in this case the Go-like potential suitably describes the statistical tendency of large conformational fluctuations around the mean structure (26, 28). In the similar way, for the genome problem, we expect that the large fluctuation of chromosomes around their mean structures can be suitably described with the Go-like potential although chromosomes do not fold into unique structures in the nucleus.

With this method we investigate the chain dynamics of chromosomes when specific interactions work in the nucleus. This method should provide a platform to examine which interactions are necessary to explain the observed structural data and which features of the genome structure are the consequences of the random Brownian motions of chromosome chains. We find necessary conditions that the model has to meet to explain the observed fluorescently visualized data of the genome structure and dynamics in a consistent way; especially the distribution of the distance between telomeres and the dynamical feature of moving telomeres are compared with the experimental data. The results presented here show that the dynamical structural modeling is a step forward to construct a unified view of the genome organization.

Methods

The bead-spring polymer model

In the present model, 16 chromosomes in the interphase haploid budding yeast are represented by 16 chains of beads and springs. Since the observed data of pairwise mean distances have resolution of a few kbp (22), we assume that each bead in the model corresponds to a 3kbp DNA segment. Each chromosome consists of 78 (Chromosome 1, or Chr1) to 806 (Chr12) beads and the total 4460 beads are considered for 16 chains.

Movement of each chromosome chain is simulated by the Langevin dynamics which is obtained by numerically solving the following equation of motion;

$$m\frac{d^2\mathbf{r}_i^\mu}{dt^2} = -\frac{\partial}{\partial \mathbf{r}_i^\mu}U - \zeta\frac{d\mathbf{r}_i^\mu}{dt} + \mathbf{w}_i^\mu, \tag{1}$$

where $\mathbf{r}_i^\mu$ is the position of the ith bead of the μth chain with μ=1–16, m is the mass of a bead, and ζ is the friction coefficient. $\mathbf{w}_i^\mu$ is the Gaussian white noise with the dispersion $< \mathbf{w}_{i\alpha}^\mu(t)\mathbf{w}_{j\beta}^\nu(t') > = 2\zeta T\delta(t-t')\delta_{ij}\delta_{\mu\nu}\delta_{\alpha\beta}$, where α and β represent the x, y or z component of the vector. Movement of chromosomes in the interphase yeast nucleus is the non-equilibrium process whose rate depends on the ATP concentration (11, 12). This energy-dependent motion, however, is random and is brought about without any detectable large motor system which may generate the coherent biased motion (12). We therefore conveniently simulate this random motion by using the effective 'temperature' T (with the unit of $k_{\rm B} = 1$) though the explicit

consideration of the non-equilibrium effects should be important in the future research. Hereafter, T is used as a unit to define interaction parameters in U.

The potential U consists of several terms;

$$U = U_{\rm LJ} + U_{\rm spring} + U_{\rm bend} + U_{\rm Go} + U_{\rm nucleus}, \tag{2}$$

where the first three terms represent the extended version of the WLC model which includes a kinkable potential as explained later in this subsection.

The term $U_{\rm LJ}$ consists of purely repulsive Lennard-Jones (LJ) type potentials representing the exclusive forces between chromatins,

$$U_{\rm LJ} = \sum_{\mu>\nu}\sum_{i,j} U_{\rm LJ}(\mu,i;\nu,j) + \sum_{\mu}\sum_{j\geq i+2} U_{\rm LJ}(\mu,i;\mu,j), \tag{3}$$

where $U_{\rm LJ}(\mu,i;\nu,j)$ has the same functional form as used by Rosa et al. (10);

$$\begin{aligned} U_{\rm LJ}(\mu,i;\nu,j) &= 4\epsilon\left(\left(\frac{a}{r_{ij}^{\mu\nu}}\right)^{12} - \left(\frac{a}{r_{ij}^{\mu\nu}}\right)^{6} + \frac{1}{4}\right), \text{ for } r_{ij}^{\mu\nu} \leq 2^{\frac{1}{6}}a, \\ &= 0, \text{ for } r_{ij}^{\mu\nu} > 2^{\frac{1}{6}}a. \end{aligned} \tag{4}$$

Here, $r_{ij}^{\mu\nu} = |\mathbf{r}_i^{\mu} - \mathbf{r}_j^{\nu}|$, and a is the thickness of the chromatin fiber, a=30nm, with the exception explained later in the subsection of Simulation. When the ordered chromatin structure is modified through fluctuation, two chromatin fibers may come closer than a (16). We adopt a mild value of $\epsilon/T = 1$ to allow two chromatin chains to approach each other.

$U_{\rm spring}$ is the potential to describe the spring between neighboring beads along the chain,

$$\begin{aligned} U_{\rm spring} &= \sum_{\mu}\sum_{i}\Big[U_{\rm r}(r_i^{\mu}) + \theta(r_i^{\mu})U_{\rm FENE}(r_i^{\mu}) \\ &+ (1-\theta(r_i^{\mu}))U_l(r_i^{\mu})\Big], \end{aligned} \tag{5}$$

where $r_i^{\mu} = |\mathbf{r}_i^{\mu} - \mathbf{r}_{i+1}^{\mu}|$, and $\theta(r_i^{\mu}) = 1$ when $r_i^{\mu} < R_0' < R_0$ and $\theta(r_i^{\mu}) = 0$ when $r_i^{\mu} \geq R_0'$. Here, $U_{\rm FENE}$ is the finitely extensible nonlinear elastic (FENE) potential (9, 10, 29),

$$U_{\rm FENE}(r_i^{\mu}) = -\frac{1}{2}kR_0^2\log\left(1 - \left(\frac{r_i^{\mu}}{R_0}\right)^2\right). \tag{6}$$

Following Rosa et al. (10), we chose R_0 to be $R_0 = 1.5\sigma_0$ with σ_0 being the typical length of the spring; By assuming the packing density of chromatin as 130bp/nm (19), $\sigma_0 \approx 3000/130 = 23$nm, so that $R_0 = 34.5$nm. k was calibrated to be $k/T = 3.5/{\sigma_0}^2$ so as to make the simulated mean distance between neighboring beads to be σ_0. The FENE potential of Eq.6 diverges at around R_0. In order to avoid the numerical instability due to this divergence, the potential is switched in Eq.5 to the milder one as $U_l = (bk/{\sigma_0}^3)(r_i^\mu)^5$ for $r_i^\mu \geq R_0' = 30$nm. For the smooth connection in this switching, we use $b = 0.42$. To keep the numerical stability, we also introduce a mild repulsive potential $U_{\rm r}$ in Eq.5 as

$$\begin{aligned} U_{\rm r}(r_i^\mu) &= 4\epsilon\left(\left(\frac{a}{r_i^\mu}\right)^2 - \left(\frac{a}{r_i^\mu}\right) + \frac{1}{4}\right), \text{ for } r_i^\mu \leq 2a, \\ &= 0, \text{ for } r_i^\mu > 2a. \end{aligned} \tag{7}$$

The potential $U_{\rm bend} = \sum_\mu \sum_i U_{\rm b}(\phi_i^\mu)$ represents the stiffness of chromatin fibers, where ϕ_i^μ is the angle between vectors $\mathbf{r}_i^\mu - \mathbf{r}_{i-1}^\mu$ and $\mathbf{r}_{i+1}^\mu - \mathbf{r}_i^\mu$. Since the local cooperative rearrangement of nucleosomes should bring about the sharp bending, or the kink of the chromosome (30, 31), we use a kinkable bending potential which saturates for $\phi > \pi/2$ (10, 32) as,

$$\begin{aligned} U_{\rm b}(\phi_i^\mu) &= k_\phi(1 - \cos\phi_i^\mu), \text{ for } 0.1 \leq \cos\phi_i^\mu \leq 1, \\ &= 0.9k_\phi, \text{ for } -1 \leq \cos\phi_i^\mu < 0.1, \end{aligned} \tag{8}$$

where the boundary value of 0.1 was chosen to fit $U_{\rm b}$ to the potential proposed in Ref.(10), but its precise value does not affect the results. The dispersion of angles should be related to the persistent length $l_{\rm p}$ and σ_0 as $<\phi^2> = 2\sigma_0/l_{\rm p}$ (33). Using the estimation of $l_{\rm p}$ ≈170-220nm for chromatins (19) and $\sigma_0 \approx 23$nm, we have $\sqrt{<\phi^2>} \approx 26 - 30^\circ$. k_ϕ was calibrated to be $k_\phi/T = 2.0$ to make the simulated results of $\sqrt{<\phi^2>}$ fall in this range.

The Go-like potential

$U_{\rm Go}$ in Eq.2 is the potential representing the tendency that distances between sites of chromosomes fluctuate around the mean distances estimated from the 3C-based method (22). We derive $\sigma_{ij}^{\mu\nu}$, the mean distance between the ith site of the μth chromosome and the jth site of the νth chromosome, by using the curve of Supplementary Figure 17 of Ref.(22) to transform the measured frequency of proximate contact into the mean distance. Using

thus derived $\sigma_{ij}^{\mu\nu}$, we have

$$U_{\mathrm{Go}} = \sum_{\mu>\nu}\sum_{i,j} U_{\mathrm{Go}}(\mu,i;\nu,j) + \sum_{\mu}\sum_{j>i} U_{\mathrm{Go}}(\mu,i;\mu,j), \tag{9}$$

with

$$U_{\mathrm{Go}}(\mu,i;\nu,j) = -\frac{\xi}{\sqrt{2\pi(c\sigma_{ij}^{\mu\nu})^2}} \exp\left(-\frac{\left(r_{ij}^{\mu\nu}-\sigma_{ij}^{\mu\nu}\right)^2}{2\left(10c\sigma_{ij}^{\mu\nu}\right)^2}\right). \tag{10}$$

Here, to simulate the large fluctuation in $r_{ij}^{\mu\nu}$, a shallow Gaussian function with $c = 0.1$ is used instead of the short-ranged Lennard-Jones type potential which is more popular in Go-like models (28). The similar Gaussian potentials for pairwise distances have been used in the structural modeling of proteins (34–36).

Since $\sigma_{ij}^{\mu\nu}$ is determined by the combined effects of the specific chromatin-chromatin interactions, the specific chromatin-nucleus interactions, and the random movement of chromosome chains, the lower value of U_{Go} indicates that the consistency among those effects is better fulfilled in the model. As compared with the effects of the chromatin-nucleus interactions and the kinetic effects represented by U_{nucleus} and $U_{\mathrm{LJ}} + U_{\mathrm{spring}} + U_{\mathrm{bend}}$ respectively, U_{Go} highlights the effects of specific chromatin-chromatin interactions which may arise, for example, from the transient formation of protein complexes that bind multiple chromosomes together. In this paper we compare the results by varying ξ in Eq.10 to see the roles of the chromatin-chromatin interactions in organization of the genome.

Constraint of nuclear structure

The nuclear structure considered in the model is illustrated in Fig.1. We use the coordinate shown in Fig.1 to explain the structure. The nucleus of interphase budding yeast is approximated by a sphere of 1μm (11, 17), so that the center of nucleus is placed on $\mathbf{r}_{\mathrm{center}} = (1000, 1000, 1000)$ in units of nanometer. The observed position of SPB is about 13nm away from the nuclear envelope (17), so that we approximate the position of SPB as $(1000, 1000, 10)$.

Centromeres of chromosomes are linked to SPB with microtubules. This linkage is represented by the potential for spring as

$$U_{\mathrm{cen}} = \sum_{\mu} \frac{h}{2}\left(\frac{l^{\mu}-l_0^{\mu}}{s}\right)^2, \tag{11}$$

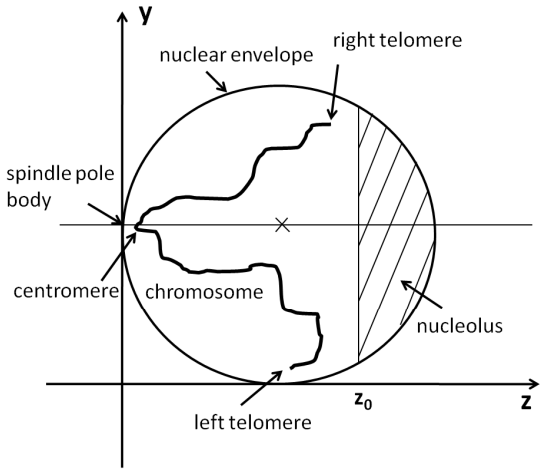

Figure 1

Figure 1: The yeast nucleus is approximated by a sphere of radius 1μm. Center of the sphere (marked by a cross) is at (1000,1000,1000) in units of nanometer in the model coordinate. Nucleolus is represented by the region of $z > z_0$. One of 16 chromosomes (Chr10) is schematically drawn in the figure. Centromere of each chromosome is linked to spindle pole body (SPB) which is a protein complex embedded in the nuclear envelope, and termini of each chromosome are left and right telomeres.

where l^μ is the length between SPB and the centromere of the μth chromosome and l_0^μ is the corresponding length defined in the model structure of Duan et al. (22). The spring constant h is chosen to allow the length variation of linker microtubules as $h/T = 0.3$ with $s = 100$nm. This small stiffness should make positions of centromeres widely spread as has been observed by the fluorescence measurement (17).

The opposite side of the nucleus is occupied by the nucleolus, in which rDNA is confined. Following Duan et al., the region from 450kbp to 1815kbp of Chr12 is regarded as rDNA (22). In order to roughly reproduce the observed distribution of rDNA (17), the nucleolus in the model is represented by the region of $z > z_0$ in Fig.1 with $z_0 = 1170$nm. The constraint imposed by nucleolus is represented by

$$\begin{aligned} U_{\rm nucleolus} &= \sum_\mu \sum_{i \notin {\rm rDNA}} U_{\rm nucl}(z_i^\mu) \\ &+ \sum_{i \in {\rm rDNA}} U_{\rm nucl}^{\rm rDNA}(z_i^{12}), \end{aligned} \tag{12}$$

with

$$\begin{aligned} U_{\rm nucl}(z_i^\mu) &= \frac{h_{\rm nucl}}{2}\left(\frac{z_i^\mu - z_0}{s}\right)^2, \text{ for } z_i^\mu \geq z_0, \\ &= 0, \text{ for } z_i^\mu < z_0\,, \text{ with } \mu \neq 12, \end{aligned} \tag{13}$$

and

$$\begin{aligned} U_{\rm nucl}^{\rm rDNA}(z_i^{12}) &= \frac{h_{\rm nucl}}{2}\left(\frac{z_i^{12} - z_0}{s}\right)^2, \text{ for } z_i^{12} < z_0, \\ &= 0, \text{ for } z_i^{12} \geq z_0, \end{aligned} \tag{14}$$

where z_i^μ is the z-component of $\mathbf{r}_i^\mu$. Since nucleolus is a soft matter, composite of nucleic acids and proteins, the boundary of nucleolus should be deformable in fluctuating environment. $h_{\rm nucl}$ is, therefore, chosen to be $h_{\rm nucl}/T = 0.2$ to allow fluctuation of a few hundred nanometers.

Chromosomes may interact with the nuclear envelope at subtelomere regions (37), and also at the actively transcribed sites (3, 38). These interactions are represented by

$$\begin{aligned} U_{\rm envelope} &= \sum_\mu \sum_i \Big(\eta(\mu,i) U_{\rm attr}(R_i^\mu) \\ &+ (1 - \eta(\mu,i)) U_{\rm rep}(R_i^\mu) \Big), \end{aligned} \tag{15}$$

where $R_i^\mu = |\mathbf{r}_i^\mu - \mathbf{r}_{\rm center}|$. Specificity of interactions are defined by η; $\eta(\mu,i) = 1$ when the interaction between the site (μ,i) and the nuclear envelope is attractive, and $\eta(\mu,i) = 0$ otherwise. We assume that the attractive interaction works only when protein complexes are formed between the site and the envelope, and hence the contact potential having the effective width of several ten nanometers is adopted. The depth of the contact should be small so as to facilitate the frequent attaching and detaching of the site to and from the envelope (11). We thus have

$$\begin{aligned} U_{\rm attr}(R_i^\mu) &= 2\epsilon \left(\left(\frac{R_i^\mu - R_0}{u - R_0} \right)^{12} - \left(\frac{R_i^\mu - R_0}{u - R_0} \right)^{6} \right), \\ &\qquad \text{for } R_i^\mu > R_0, \\ &= 0, \text{ for } R_i^\mu \le R_0, \end{aligned} \tag{16}$$

and

$$\begin{aligned} U_{\rm repul}(R_i^\mu) &= 2\epsilon \left(\frac{R_i^\mu - R_0}{u - R_0} \right)^{12}, \text{ for } R_i^\mu > R_0, \\ &= 0, \text{ for } R_i^\mu \le R_0, \end{aligned} \tag{17}$$

with $R_0 = 800$nm and $u = 1000$nm.

We compare the results of four different models of $\eta(\mu,i)$ to examine which chromatin-envelope interactions are important to explain the observed data. In Model 1 and Model 2, no specificity is assumed: In Model 1, $\eta(\mu,i) = 0$ for all sites (μ,i), and in Model 2, $\eta = 1$ for all sites. Heterogeneous interactions, on the other hand, are assumed in Model 3 and Model 4. In Model 3, $\eta = 1$ at the telomere sites and also at the sites in the rDNA region, but $\eta = 0$ at other sites. Here, depending on the observed telomere length (39), we regard from one to consecutive four sites at the end of each chromosome chain as a telomere. In Model 4, $\eta = 1$ is assumed at all telomere sites, at the rDNA sites, and also at the sites which are in contact with the nuclear envelope in the Duan et al. structure (22), and $\eta = 0$ at other sites.

By summing Eqs.11, 12, and 15, we obtain the potential for the nuclear constraint, $U_{\rm nucleus}$, as

$$U_{\rm nucleus} = U_{\rm cen} + U_{\rm nucleolus} + U_{\rm envelope}, \tag{18}$$

which is the last term of Eq.2.

Simulation

First, the model structure proposed by Duan et al. (22) was modified to lower the potential energy U. Thus obtained relaxed structure was used as the initial structure of simulation. When the distance between two sites was smaller than 30nm in this initial structure, we used that distance as a in Eq.4.

It should be noted that the Duan et al. structure is one of many possible structures which satisfy the constraints arising from $\{\sigma_{ij}^{\mu\nu}\}$ to certain extent. We will see in Results section that the Langevin dynamics indeed generates many structures which deviate from the Duan et al. structure.

Starting from thus obtained initial structure, the Langevin dynamics of the genome was followed numerically. Using the unit of $m = T = 1$ in Eq.1, time t has the dimension of length L and the friction constant ζ has the dimension of L^{-1} in the simulation. With this unit, Eq.1 was discretized with the interval $\Delta t = 0.01$ for one step of the simulation and ζ was set to be $\zeta = 10^{-5}$ to allow efficient sampling in the allowed computation time. The first 10^4 steps were used to equilibrate the system, and then the subsequent 5×10^4 steps or 1.1×10^5 steps were sampled for obtaining the statistical data. 10 independent runs with the different random number realization were performed and the distributions of chromosomes were derived from this ensemble of data.

Results

Large fluctuation of the genome structure

Shown in Fig.2 is a snapshot of the genome structure calculated with the Langevin dynamics. The Langevin dynamics simulates the fluctuating motion of chromosome chains under the influence of the potential energy U. Through the Langevin dynamics, the genome structure is largely deformed from the initial Duan et al. structure. This deformation is due to the difference in the way to model nucleolus: In the Duan et al. structure, the nucleolus was modeled as a small sphere of volume $0.11\mu\text{m}^3$ (22), while in the fluorescence data, the estimated volume of the region containing 50% of observed locus positions of rDNA was $0.53\mu\text{m}^3$ (17). If we assume that rDNA is confined in the nucleolus, apparently the nucleolus should be modeled to have a larger volume than in the Duan et al. structure. In the present simulation the nucleolus is modeled with a more realistic size, so

that rDNA expands in the larger volume of the nucleolus and the whole genome is pushed by the effective pressure of the bulky nucleolus toward the SPB side from the initial structure. Through such deformation during the equilibration process, the simulated chromosomes distribute in the nucleus as shown in Fig.2.

After the equilibration process, each part of the genome continues to fluctuate largely. In Fig.3, motions of centromeres (Fig.3a), distribution of rDNA (Fig.3b), and motions of telomeres (Fig.3c) and genes (Fig.3d) are shown. The trajectories of each gene and telomere traverse over a half of the radius of nucleus, showing the large fluctuation of the genome structure. These trajectories, however, are not largely overlapped with each other. The amplitude of chromosome fluctuation is not large enough to homogenize nucleus completely but parts of chromosomes are separated to form 'territories' as has been pointed out by using the fluorescence data (17). See also the supplementary movie S1 to grasp the feeling of the chromosome dynamics.

Specific chromatin-chromatin interactions

We compare the results by varying the strength ξ of the Go-like potential from $\xi/T = 1$ to 100. ξ/T represents the degree of how strongly the chromosome chains are constrained around their mean structures by the chromatin-chromatin interactions in the model. With $\xi/T = 1$, we can expect that the chromatin-chromatin interactions represented by the Go-like potential are so weak that the other effects such as effects of fluctuating motions and constraints of the nuclear structure dominate dynamics of the system, while with $\xi/T = 100$ the chromatin-chromatin interactions dominate dynamics to give rise to a less flexible genome conformation.

Plotted in Fig.4 are the observed (40) and simulated distributions of distances between telomeres. Each distribution has a large width showing the large amplitude motion of telomeres. The precise form of distribution, however, depends on the pair of telomeres examined, showing the heterogeneous fluctuations in the genome: Distance between the left telomere of Chr3 (3L) and the right telomere of Chr3 (3R) tends to be small, but the larger 6L-6R distance is observed and the 5L-5R and 14L-14R distances have further large variation.

The simulated results can semi-quantitatively reproduce the observed data when the strength ξ is appropriately chosen: The distributions simulated with $\xi/T = 10$ can reasonably fit the experimental data for the 3L-3R, 6L-6R, 14L-14R, and 6L-14L distances, while the distribution simulated with

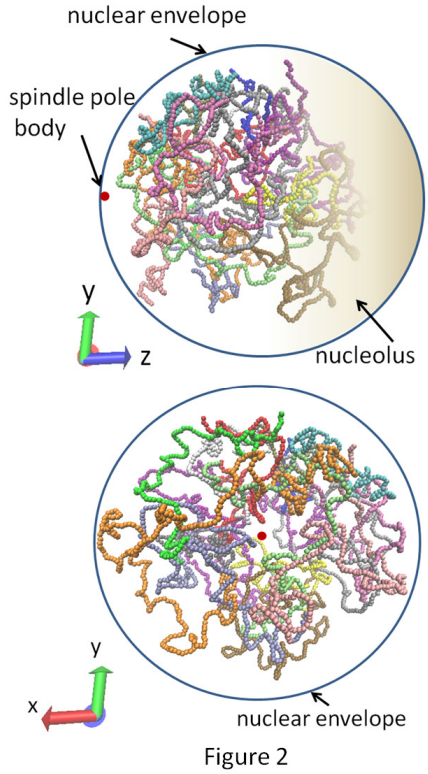

Figure 2: A snapshot of the simulated genome structure. Structures of 16 chromosomes are shown by different colors; Chr1 (blue), Chr2 (red), Chr3 (gray), Chr4 (orange), Chr5 (yellow), Chr6 (tan), Chr7 (silver), Chr8 (green), Chr9 (white), Chr10 (pink), Chr11 (cyan), Chr12 (purple), Chr13 (lime), Chr14 (mauve), Chr15 (ochre), and Chr16 (iceblue). SPB is represented by a red dot. The shaded region is nucleolus. Top: a snapshot drawn from the angle similar to that in Figure 1. Bottom: the same structure viewed from the SPB side. Simulated with $\xi/T = 10$ and Model 3.

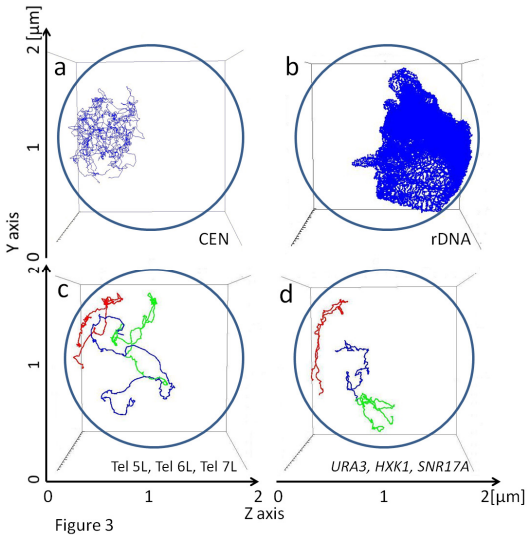

Figure 3: Examples of simulated trajectories of parts of the genome. (*a*) Trajectories of 16 centromeres are superposed, (*b*) traces of the rDNA region moving during the simulation, (*c*) trajectories of the left telomere of Chr5 (5L, blue), the left telomere of Chr6 (6L, red), and the left telomere of Chr7 (7L, green), and (*d*) trajectories of genes, *ura3* (blue), *hxk1* (red), and *snr17a* (green). *hxk1* is located near the telomere of Chr6, *ura3* is near the centromere of Chr5, and *snr17a* is in between centromere and telomere of Chr15. Four spheres from *a* to *d* are viewed from the same angle. In *a*, centromeres move around SPB, while in *b*, rDNA spreads inside the nucleolus. In *c*, 6L moves near the nuclear envelope, 5L is bound and unbound to and from the nuclear envelope, and 7L moves more freely. in *d*, genes move separately to form 'gene territories'. Trajectories of 1.1×10^5 steps simulated with $\xi/T = 10$ and Model 3 are shown.

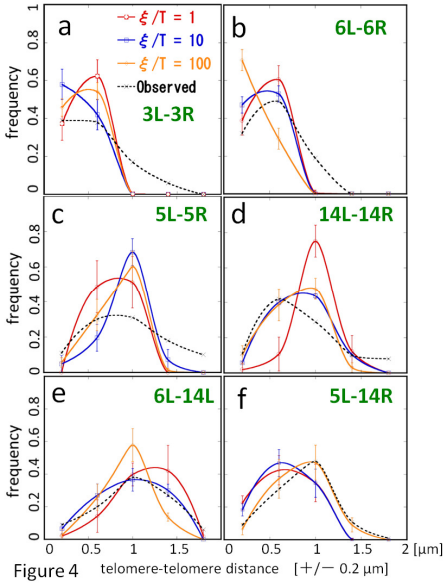

Figure 4: Dependence of telomere-telomere distance distributions on the strength of chromatin-chromatin interactions. Data simulated with $\xi/T = 1$ (red), $\xi/T = 10$ (blue), and $\xi/T = 100$ (orange) are compared with the data observed with the fluorescently labeled proteins (black dotted line) (40). Distributions between (*a*) 3L and 3R, (*b*) 6L and 6R, (*c*) 5L and 5R, (*d*) 14L and 14R, (*e*) 6L and 14L, and (*f*) 5L and 14R. Points obtained by binning data over $\pm 0.2\mu$m are connected by smooth lines. Error bars are standard deviations of trajectories simulated for 5×10^4 steps with Model 3.

$\xi/T = 1$ fails to fit the 14L-14R distance. The distribution simulated with $\xi/T = 100$ can fit the observed 5L-14R distance very well but fails to explain the 6L-6R distance. The 5L-5R distance can not be fitted well by all the simulated results, which may be due to the limited sampling timesteps of trajectories. As shown in Fig.3c (and will be also shown in Fig.6a), 5L binds and unbinds to and from the nuclear envelope during the simulation, so that it should need the longer trajectories to sample enough data for the equilibrium distribution of the 5L-5R distance. Such large fluctuation of the position of 5L in the fluorescence data has been also reported (13).

We should note that the Go-like potential represents the consistency among the chromatin-chromatin interactions and other effects in nucleus and does not directly represent the physical interactions working through the formation of complexes of proteins that bind multiple chromosomes. The necessity of strong ($\xi/T = 100$) or moderately strong ($\xi/T = 10$) Go-like potential for modeling the genome to reproduce the observed telomere-telomere distributions, however, strongly suggests that chromosomes are not the non-specific polymer chains but the specific chromatin-chromatin interactions play important roles in organizing large but characteristic motions of chromosomes.

Heterogeneous interactions between chromosome and nuclear envelope

Also compared with the fluorescence data are four different models of attractive chromatin-envelope interactions represented by different distributions of $\{\eta(\mu, i)\}$. Here, $\eta(\mu, i)$ is an index representing whether there is an attractive interaction between the ith site of the μth chain and the nuclear envelope (η=1) or there is no such an interaction between them (η=0); Model 1 ($\eta = 0$ for all sites), Model 2 ($\eta = 1$ for all sites), Model 3 ($\eta = 1$ for telomere and rDNA sites, and $\eta = 0$ for other sites), and Model 4 ($\eta = 1$ for telomere and rDNA sites and sites near the nuclear envelope in the Duan et al. structure, and $\eta = 0$ for other sites) were tested.

Since the regions of rDNA and the regions near telomeres are often found around the nuclear envelope in the fluorescence observations, the attractive interactions have been expected between rDNA and nuclear envelope (41) and also between telomeres and nuclear envelope (37). It is natural, therefore, that Model 1 does not explain the observed distributions (40) as shown in Fig.5. In Fig.5, we also find that Model 2 is inconsistent with the observed data, which implies that not every site but only the selected sites

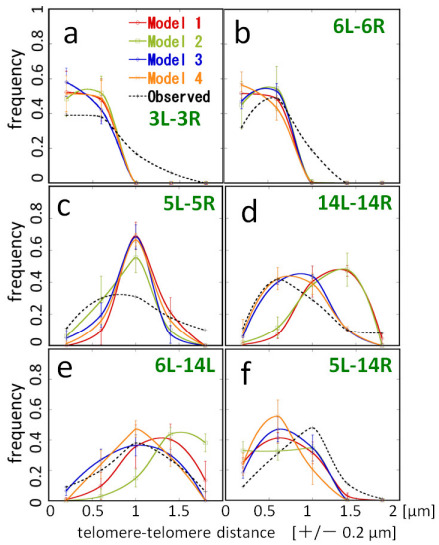

Figure 5

Figure 5: Dependence of telomere-telomere distance distributions on the interactions between chromatins and the nuclear envelope. Data simulated with Model 1 (red), Model 2 (green), Model 3 (blue), and Model 4 (orange) are compared with the data observed with the fluorescently labeled proteins (black dotted line) (40). Distributions between (*a*) 3L and 3R, (*b*) 6L and 6R, (*c*) 5L and 5R, (*d*) 14L and 14R, (*e*) 6L and 14L, and (*f*) 5L and 14R. Points obtained by binning data over $\pm 0.2\mu$m are connected by smooth lines. Error bars are standard deviations of trajectories simulated for 5×10^4 steps with $\xi/T = 10$.

should form attractive protein complexes with the nuclear envelope.

Model 3 and Model 4 explain the observed data to a similar extent. Although Model 4 has the additional sites attractive to the nuclear envelope, the number of such sites is not large (11 sites in Chr1, 4 sites in Chr4, and 2 sites in Chr6), so that the effects are small in the resolution of Fig.5. It would be interesting to further examine whether the more precise experimental measurement can distinguish the difference in distribution of $\{\eta(\mu, i)\}$.

Diffusive movement of telomeres

Interesting features of the simulated data are on the dynamical movement of chromosomes. Shown in Fig.6a is the distance between telomeres and

the nuclear envelope. We can find that the telomere 5R is not bound to the nuclear envelope but 5L is transiently bound and unbound to and from the nuclear envelope. 6R and 6L are bound to the nuclear envelope and distances between those sites and the nuclear envelope oscillate within a few hundred nanometers. These simulated features are consistent with the observed dynamical features of 5R, 5L, 6R, and 6L (13).

In order to compare the simulated data with the observed one, we define the radial mean square displacement $msd_{\rm R}$ of the site (μ, i) during the time steps t as

$$msd_{\rm R}(t) = \left\langle (R_i^\mu(t+\tau) - R_i^\mu(\tau))^2 \right\rangle_\tau , \tag{19}$$

where $\langle \dots \rangle_\tau$ is the average over τ for the trajectory examined. In Fig.6b $msd_{\rm R}(t)$ is plotted as a function of t. With this plot we can clearly see the difference between the results with $\xi/T = 10$ and those with $\xi/T = 100$: Fig.6b shows that $msd_{\rm R}$ of 5L and 5R with $\xi/T = 10$ is roughly proportional to t indicating that their motion is diffusive. With $\xi/T = 100$, on the other hand, $msd_{\rm R}$ shows a complex pattern with the oscillatory behaviors, which indicates the elastic features of the genome structure arising from the strong chromatin-chromatin interactions. The experimental data shows that $msd_{\rm R}$ is diffusive without exhibiting an oscillatory behavior (13). We should conclude, therefore, that in yeast genome the constraints due to the chromatin-chromatin interactions are not so strong as to generate the elastic stiffness but are at the moderate level to keep fluidity of the genome structure in nucleus. By comparing the slope of $msd_{\rm R}(t)$ for $\xi/T = 10$ with the experimental data (13), the suggested length of a step in the simulation is 2.63×10^{-3} sec/step when the 5L data is used and 4.47×10^{-3}sec/step when the 5R data is used. With this estimation of order of time scale, the length of trajectories shown in Fig.6a corresponds to about 290-490 seconds.

Discussion

In this paper a method was developed for modeling dynamical 3D organization of genome of interphase budding yeast. The modeled genome exhibits a large structural fluctuation with the diffusive motion of chromosomes. In spite of such intense fluctuation, the simulated movement of chromosomes is not completely random but is subjected to both the specific chromatin-chromatin interactions and the heterogeneous chromatin-envelope interactions: By suitably taking into account the information obtained from the

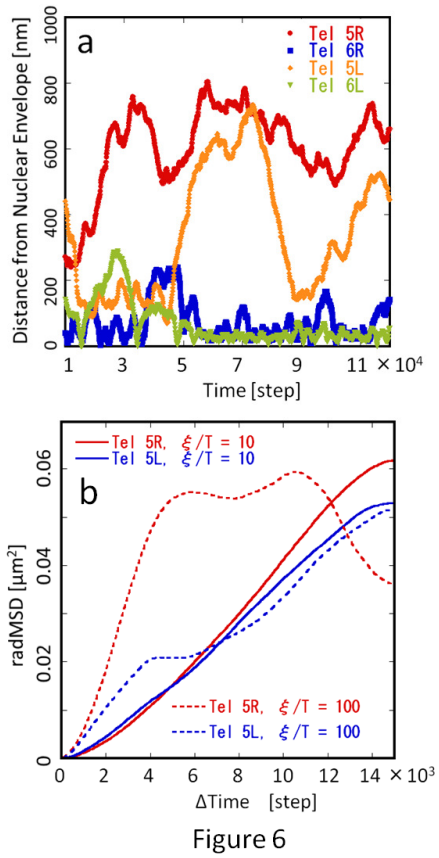

Figure 6

Figure 6: The simulated telomere dynamics. (a) Temporal change of the simulated distance between telomeres and the nuclear envelope is shown for 5R, 5L, 6R, and 6L. Simulated with $\xi/T = 10$ and Model 3. (b) $msd_{\mathrm{R}}(t)$ of 5R and 5L are plotted as functions of the number of time steps t. Simulated data for ξ/T=10 (real lines) and 100 (dotted lines) with Model 3 are shown for 5L (blue) and 5R (red). By comparing the slope of simulated $msd_{\mathrm{R}}(t)$ with the observed one (13), it is suggested that 11×10^4 steps in simulation roughly correspond to 290-490 seconds (see text for this estimation).

3C-based method, the model explained the fluorescence data for the distribution and movement of each part of the genome.

We should note that an important point to be improved in the present modeling is on the treatment of nucleolus. In the present model, nucleolus was considered as a force field acting on chromosomes. Nucleolus, however, is a complex of ribosomal DNA, RNA, and related proteins and should behave as a deformable substance. Moving chromosomes should push the nucleolus to deform it to a concave form and the deformed nucleolus should then apply forces on chromosomes in a different way from that considered in the present model. By treating nucleolus in a more realistic way, we may be able to construct a model which should have more quantitative prediction capability.

With such improvement, we will be able to explore many challenging problems. For example, genes actively transcribed may be anchored around the nuclear pore (3, 38). The attractive interaction between the gene at site (μ, i) and the nuclear envelope is represented by $\eta(\mu, i)$ in the present model, so that the model may predict how the gene expression pattern and the observation on the movement of that gene is correlated. Also interesting is on the protein factors involved in the complex to anchor chromosomes to the nuclear envelope. Since the formation of complexes may bring about the sequestration of such factors, the model may predict the spatiotemporal pattern of confinement and release of those factors. Further interesting aspects are on the dynamical movement of chromosomes. We may analyze it, for example, by decomposing the movement into principal components. It would be intriguing to see whether there is a correlation between those components and principal modes of the temporal variation of gene expression.

Also interesting is to see the effects of heterogeneous rigidity of the chromosome chains. It has been shown that the loci of induced DSB are dislocated to the nuclear periphery to form the repair domain (42, 43). The localized modulation of stiffness parameters such as k in Eq.6 or k_ϕ in Eq.8 should represent DSB at the corresponding loci in the model, and hence it should be possible to examine whether the DSB loci move owing to interactions mediated by specific proteins or through the biased diffusion in the organized genome structure.

In this way the efforts to construct the computational model of dynamical 3D genome organization should lead to a unified view of the genome structure, dynamics, and function, so that it should open a new field of interacting computational and experimental biophysics of structural genetics.

Supplementary Material

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

Supplementary Video S1

A typical example of the simulated motion of 16 chromosomes in the interphase yeast nucleus. Shown is the trajectory from the 27,500th to 60,000th step simulated with $\xi/T = 10$ and Model 3. 16 chromosomes are distinguished by different colors; Chr1 (blue), Chr2 (red), Chr3 (gray), Chr4 (orange), Chr5 (yellow), Chr6 (tan), Chr7 (silver), Chr8 (green), Chr9 (white), Chr10 (pink), Chr11 (cyan), Chr12 (purple), Chr13 (lime), Chr14 (mauve), Chr15 (ochre), and Chr16 (iceblue). The purple chain poking out to the right is Chr12 which contains rDNA.

References

1. Misteli, T., 2007. Beyond the sequence: cellular organization of genome function. *Cell* 128:787–800.

2. Miele, A., and J. Dekker, 2008. Long-range chromosomal interactions and gene regulation. *Mol. Biosyst.* 4:1046–1057.

3. Dillon, N., 2008. The impact of gene location in the nucleus on transcriptional regulation. *Dev. Cell* 15:182–186.

4. Zimmer, D., and E. Fabre, 2011. Principles of chromosomal organization: lessons from yeast. *J. Cell. Biol.* 192:723–733.

5. Lanctot, C., T. Cheutin, ..., T. Cremer, 2007. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8:104–115.

6. Lieberman-Aiden, E., N. L. van Berkum, ..., J. Dekker, 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–293.

7. Grosberg, A., Y. Rabin, ..., A. Neer, 1993. Crumpled globule model of the three-dimensional structure of DNA *Europhys. Lett.* 23:373–378.

8. Mirny, L. A., 2011. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* 19:37–51.

9. Rosa, A., and R. Everaers, 2008. Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* 4:e1000153.

10. Rosa, A., N. B. Becker, and R. Everaers, 2010. Looping probabilities in model interphase chromosomes. *Biophys. J.* 98:2410–2419.

11. Heun, P., T. Laroche, ..., S. M. Gasser, 2001. Chromosome dynamics in the yeast interphase nucleus. *Science* 294:2181–2186.

12. Gasser, S. M., 2002. Visualizing chromatin dynamics in interphase nuclei. *Science* 296:1412–1416.

13. Hediger, F., F. R. Neumann, ..., S. M. Gasser, 2002. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr. Biol.* 12:2076–2089.

14. Schober, H., V. Kalck, ..., S. M. Gasser, 2008. Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Res.* 18:261-271.

15. Therizols, P., T. Duong, ..., E. Fabre, 2010. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. USA* 107:2025–2030.

16. Maeshima, K., S. Hihara, and M. Eltsov, 2010. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr. Opn. Cell Biol.* 22:291–297.

17. Berger, A. B., G. G. Cabal, ..., C. Zimmer, 2008. High-resolution statistical mapping reveals gene territories in live yeast. *Nat. Methods* 5:1031–1037.

18. Ferrai, C., I. J. de Castro, ..., A. Pombo, 2010. Gene positioning. *Cold Spring Harbor Perspect. Biol.* 2:a000588.

19. Bystricky, K., P. Heun, ..., S. M. Gasser, 2004. Long range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. USA* 101:16495–16500.

20. Taddei, A., H. Schober, and S. M. Gasser, 2010. The budding yeast nucleus. *Cold Spring Harbor Perspect. Biol.* 2:a000612.

21. Torres-Rosell, J., I. Sunjevaric, ..., M. Lisby, 2007. The Smc5-Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. *Nat. Cell Biol.* 9:923–931.

22. Duan, Z., M. Andronescu, ..., W. S. Noble, 2010. A three-dimensional model of the yeast genome. *Nature* 465:363–367.

23. Tanizawa, H., O. Iwasaki, ..., K. Noma, 2010. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.* 38:8164–8177.

24. Lisby, M., U.H. Mortensen, and R. Rothstein, 2003. Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. *Nat. Cell Biol.* 5:572–577.

25. Marti-Renom, M. A., and L. A. Mirny, 2011. Bridging the resolution gap in structural modeling of 3D genome organization. *PLoS Comput. Biol.* 7:e1002125.

26. Onuchic, J. N., and P. G. Wolynes, 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.

27. Taketomi, A., Y. Ueda, and N. Go, 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. 1. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7:445–459.

28. Clementi, C., H. Nymeyer, and J. N. Onuchic, 2000. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.

29. Kremer, K., and G. S. Grest, 1990. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys.* 92:5057–5086.

30. Worcel, A., S. Strogatz, and D. Riley, 1981. Structure of chromatin and the linking number of DNA. *Proc. Natl. Acad. Sci. USA.* 78:1461–1465.

31. Lankas, F., R. Lavery, and J. H. Maddocks. 2006. Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure* 14:1527–1534.

32. Wiggins, P. A., R. Phillips, and P. C. Nelson, 2005. Exact theory of kinkable elastic polymers. *Phys. Rev. E* 71:021909.

33. Hagerman, P. J., 1988. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* 17:265–286.

34. Goldstein, R. A., Z. A. Luthey-Schulten, and P. G. Wolynes, Optimal protein-folding codes from spin-glass theory. 1992. *Proc. Natl. Acad. Sci. USA* 89:4918–4922.

35. Sasai, M., 1995. Conformation, energy, and folding ability of selected amino acid sequences. *Proc. Natl. Acad. Sci. USA* 92:8438–8442.

36. Sasaki, T. N., H. Cetin, and M. Sasai, 2008. A coarse-grained Langevin molecular dynamics approach to de novo protein structure prediction. *Biochem. Biophys. Res. Commun.* 369:500–506.

37. Taddei, A., G. V. Houwe, ..., S. M. Gasser, 2009. The functional importance of telomere clustering: global changes in gene expression result from SIR factor dispersion. *Genome Res.* 19:611–625.

38. Taddei, A., G. V. Houwe, ..., S. M. Gasser, 2006. Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature* 441:774–778.

39. Saccharomyces Genome Database, http://www.yeastgenome.org/

40. Bystricky, K., T. Laroche, ..., S. M. Gasser, 2005. Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *J. Cell Biol.* 168:375–387.

41. Mekhail, K., J. Seebacher, ..., D. Moazed, 2008. Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature* 456:667–670.

42. Therizols, P., C. Fairhead, ..., E. Fabre, 2006. Telomere tethering at the nuclear periphery is essential for efficient DNA double strand break repair in subtelomeric region. *J. Cell Biol.* 172:189–199.

43. Nagai, S., K. Dubrana, ..., N.J. Krogan, 2008. Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. *Science* 322:597–602.