Physical tethering and volume exclusion determine higher-order genome organization in budding yeast


*Genome Res.* published online May 22, 2012
Access the most recent version at doi:10.1101/gr.129437.111

Supplemental Material

http://genome.cshlp.org/content/suppl/2012/04/20/gr.129437.111.DC1.html

Published online May 22, 2012 in advance of the print journal.

Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Genome Research* go to: http://genome.cshlp.org/subscriptions

© 2012, Published by Cold Spring Harbor Laboratory Press
Physical tethering and volume exclusion determine higher-order genome organization in budding yeast

Harianto Tjong,1 Ke Gong,1 Lin Chen,1,2,3 and Frank Alber1,4

1Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA; 2Department of Chemistry and Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA

In this paper we show that tethering of heterochromatic regions to nuclear landmarks and random encounters of chromosomes in the confined nuclear volume are sufficient to explain the higher-order organization of the budding yeast genome. We have quantitatively characterized the contact patterns and nuclear territories that emerge when chromosomes are allowed to behave as constrained but otherwise randomly configured flexible polymer chains in the nucleus. Remarkably, this constrained random encounter model explains in a statistical manner the experimental hallmarks of the S. cerevisiae genome organization, including (1) the folding patterns of individual chromosomes; (2) the highly enriched interactions between specific chromatin regions and chromosomes; (3) the emergence, shape, and position of gene territories; (4) the mean distances between pairs of telomeres; and (5) even the co-location of functionally related gene loci, including early replication start sites and tRNA genes. Therefore, most aspects of the yeast genome organization can be explained without calling on biochemically mediated chromatin interactions. Such interactions may modulate the pre-existing propensity for co-localization but seem not to be the cause for the observed higher-order organization. The fact that geometrical constraints alone yield a highly organized genome structure, on which different functional elements are specifically distributed, has strong implications for the folding principles of the genome and the evolution of its function.

[Supplemental material is available for this article.]
can not explain many of the specific patterns observed in experiments (Rosa et al. 2010).

To fairly assess the principles of chromosome folding and the possible role of molecular interactions in establishing nuclear order, we must first examine the genome structure that arises when chromosomes are tethered but otherwise randomly configured in the confinement of the nuclear environment. Previous work points toward an important role of nuclear constraints and relative chromosome arm lengths in genome organization (Berger et al. 2008; Taddel et al. 2010; Zimmer and Fabre 2011) as shown for the dynamic relationship of subtelomeric regions (Therizols et al. 2010). However, it remains to be seen if entirely random configurations of tethered chromosomes are sufficient to reproduce in a statistical manner all the available quantitative data about the yeast genome organization and gene loci interactions, including the highly structured contact frequency maps from genome-wide conformation capture experiments (Dekker et al. 2002; Duan et al. 2010), the distribution of gene territories from fluorescence imaging (Berger et al. 2008; Therizols et al. 2010; Zimmer and Fabre 2011), and the clustering of replication start sites as well as tRNA genes.

Our findings demonstrate that purely random configurations of tethered chromosomes do indeed reproduce in a statistical manner a wide range of data related to genome structure: genome-wide chromatin interaction frequencies; the emergence, shape, and location of telomeres; and even the spatial clustering of functionally related chromosome regions such as early replication start sites and tRNA gene loci. Specific molecular interactions between chromatin regions, although possible, are not required to explain the available experimental data on the higher-order genome organization. Moreover, the large structural variability among individual cell’s genome configurations indicates that no single average genome structure can adequately reflect the wide range of structural features relevant to a population of cells.

### Results

#### Population modeling for determining the three-dimensional organization of the genome

To address the challenge of representing highly variable genome structures, we construct a large population of three-dimensional (3D) genome structures, which represent a spectrum of all possible chromosome configurations, and interpret the result in terms of probabilities of a sample drawn from a population of heterogeneous structures (Methods).

All chromosomes are modeled as random configurations that are subject to the following constraints: (1) All chromosomes are confined in the nucleus; (2) all the centromeres are attached to the SPB through microtubules; (3) all the telomeres are located near the nuclear periphery; and (4) the nucleolus is inaccessible to chromosomes, except for those regions containing rDNA repeats (Methods) (Table 1; Fig. 1).

To generate a population of genome structures, we defined an optimization problem (Methods). In order to sample a representative set of all possible structures, we created a sample of 200,000 independently optimized genome structures, hereafter referred to as the structure population. We also generated a control population with an identical setup but without imposing any landmark constraints (Methods), referred to as the random control. We also calculated a structure population for a nucleus containing only a single chromosome, constrained in a manner identical to the full simulation. We refer to this population as the single chromosome population.

#### Probabilistic analysis of chromosome structural features

In the following sections, we analyze the spatial properties of the structure population in terms of several statistical quantities: (1) chromosome territory locations, (2) chromosome and gene loci interaction frequencies, (3) locus localization probabilities, (4) telomere distance distributions, (5) physical proximity of functionally associated genomic loci including early and late replication origins, and tRNA genes. Each property of the simulated structure population will be compared with available experimental data.

#### Chromosome territories as a result of constrained random encounters

We first ask to what extent the landmark constraints lead to preferred chromosome locations. We calculate the probability that each chromosome occupies any given region of the nucleus (i.e., the localization probability density (LPD) of a chromosome) (Supplemental Material). Based on the LPD, it is evident that all the chromosomes have preferred regions. Smaller chromosomes (e.g., chromosome 1 in Fig. 2) reside preferentially around the central axis, near the SPB. Interestingly medium-sized chromosomes are more likely to reside away from the central axis (e.g., chromosome 8 in Fig. 2A,B; Supplemental Fig. 1), while for large chromosomes

---

### Table 1. Functional forms of the restraints in the scoring function

<table>
<thead>
<tr>
<th>Restraint type</th>
<th>Functional form</th>
<th>( \mu )</th>
<th>( d ) (nm)</th>
<th>Bead ( i )</th>
<th>( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U_{nu} ) Nuclear envelope restraint</td>
<td>( u^{ab} )</td>
<td>(0,0,0)</td>
<td>1000</td>
<td>All beads</td>
<td>1</td>
</tr>
<tr>
<td>( U_{nu} ) Nucleolus localization restraint</td>
<td>( u^{ab} )</td>
<td>(−1200,0,0)</td>
<td>1600</td>
<td>( \forall i \in \delta )</td>
<td>1</td>
</tr>
<tr>
<td>( U_{nu} ) Nucleolus excluded volume restraint</td>
<td>( u^{ab} )</td>
<td>(−1200,0,0)</td>
<td>1600</td>
<td>( \forall i \in \beta )</td>
<td>1</td>
</tr>
<tr>
<td>( U_{cn} ) Centromere localization restraint</td>
<td>( u^{ab} )</td>
<td>(−700,0,0)</td>
<td>300</td>
<td>( \forall i \in \gamma )</td>
<td>1</td>
</tr>
<tr>
<td>( U_{tl} ) Telomere localization restraint</td>
<td>( u^{ab} )</td>
<td>(0,0,0)</td>
<td>950</td>
<td>( \forall i \in \gamma )</td>
<td>1</td>
</tr>
<tr>
<td>( U_{cb} ) Chromatin chain bond restraint</td>
<td>( u^{ab} )</td>
<td>( r_{i+1} )</td>
<td>30</td>
<td>( i \in {1..N−1}, j &gt; i )</td>
<td>1</td>
</tr>
<tr>
<td>( U_{cv} ) Chromatin excluded volume restraint</td>
<td>( u^{ab} )</td>
<td>( r_{i} )</td>
<td>30</td>
<td>( i \in {1..N−1}, j &gt; i )</td>
<td>1</td>
</tr>
</tbody>
</table>

All restraints are expressed as harmonic functions \( (\beta^{2}) \), as well as harmonic upper \( (\beta^{2}) \) and lower bounds \( (\beta^{2}) \), respectively: \( \beta'(r, \mu) = \frac{1}{2} k (r_i - \mu - d)^2 \), \( \beta'(r, \mu) = \frac{1}{2} k (r_i - \mu - d)^2 \), for \( r_i - \mu > d \), and \( \beta'(r, \mu) = \frac{1}{2} k (r_i - \mu - d)^2 \), for \( r_i - \mu < d \), where \( \mu \in \mathbb{R}^3 \) is the coordinate vector of a reference point, \( d \) is reference distance, \( k \) is harmonic constant, and \( N \) is the total number of beads in a model. We also define several subsets of beads that share certain properties. More specifically, \( \alpha \) is the set of beads assigned to the last bead of every chain, \( \beta \) is the set of beads assigned to centromeric regions, \( \gamma \) is the set of beads assigned to telomeric regions, and \( \delta \) is the set of all beads flanking rDNA repeat regions.
The LPD is highest in the central region of the nucleus again along the central axis. We then ask what factors are responsible for the chromosomes' preferred locations. For each chromosome, we calculate a new structure population for a nucleus containing only a single chromosome but otherwise constrained in a manner identical to the full simulation (i.e., the single chromosome population) (Fig. 2C). Comparing the two structure populations reveals great differences for each chromosome location (Fig. 2D). For example, in the full simulation, large chromosomes reside substantially farther from the SPB region toward the nucleolus than would be expected based on chromosome tethering alone. The differences are caused by a volume exclusion effect: Because of tethering, the chromosomes must compete for the limited space around the SPB. Smaller chromosomes are naturally more restricted to regions closer to the SPB, which in turn tends to exclude parts of larger chromosomes from these regions. For smaller chromosomes, the opposite effect is observed; in the full simulation, they exhibit an increased probability density around the SPB. Importantly, due to the volume exclusion effect, the preferred location of a chromosome is not defined by tethering alone but also depends on the total number and lengths of all other chromosomes in the nucleus.

Genome-wide chromosome contact patterns

Next, we measure how often any two chromosome chains come into contact with each other over the entire structure population. Interestingly, most chromosomes show distinct preferences for interacting with certain others. For instance, chromosome 1 has a significantly higher chance of interacting with chromosomes 3 and 6 than with any other chromosome. Its interactions with the large chromosomes 4, 7, and 12 are substantially depleted (Fig. 3A). Strikingly, almost identical chromosome interaction preferences are observed in an independent genome-wide chromosome conformation capture experiment (Fig. 3A; Supplemental Fig. 2A; Duan et al. 2010). Pearson's correlation between the chromosome-pair contact frequencies in our structure population and those detected in the experiment is 0.94 ($P < 10^{-15}$). In the random control, the contact frequencies do not display any significant chromosome-pair contact preferences (Pearson's correlation coefficient of 0.57) (Supplemental Fig. 2B).

Next, we compare contact frequencies for all possible pairings of the 32 chromosome arms (Fig. 3B,C). It is evident that some pairs of chromosome arms have a greater propensity to interact than others. In particular, chromosome arms with $<500$ kb (chromosomes 1, 3, 5, 6, 8, and 9) are more likely to interact with each other than longer arms. For instance, the short arm of chromosome 1R is almost eight times more likely to interact with the short arm of chromosome 3L than with the long arm of 4R. Also these observations are in almost complete agreement with the conformation capture experiments (Pearson's correlation coefficient of 0.93, $P < 10^{-15}$) (Fig. 3C,D; Duan et al. 2010).

Finally, when chromatin contacts are analyzed at a resolution of 32 kb, the contact frequency heat map of the structure population shows highly organized cross-shaped patterns (Fig. 3E).
in this situation when all chromosomes are present (see other chromosomes in the nucleus. The density distribution is significantly
landmark constraints, but structures are generated without the presence of
the ‘‘single chromosome population.’’ The chromosome is subject to all
chromosome arms are very low (blocks a in Fig. 4A).

Similar contact patterns have been reported in 3C conformation capture experiments and have been explained by a particular Rabl-like style of chromosome folding (Dekker et al. 2002). The hypothesis is that regions on opposite sides of the centromere are folded toward each other, possibly indicating the existence of a biochemical attraction between loci (Fig. 4B). However, the strong agreement between the experimental contact frequency maps and our structure population demonstrates that such contact patterns are not necessarily caused by specific biochemically mediated interations between subcentromeric regions. An equally possible explanation is that they represent purely random encounters of constrained chromosome chains.

It remains to be determined which factors are most responsible for the folding. In the ‘‘single chromosome population,’’ the cross-shaped intrachromosomal contact pattern is lost; the contact frequency map is similar to the random control (Fig. 4A, bottom panels). Therefore the particular folding pattern illustrated in Figure 4B is caused by a volume exclusion effect as a result of the presence of all 16 chromosomes. The competition among all centromere-tethered strands for the limited space around the SPB naturally leads to the style of folding described by experiments, and this folding is the proximate cause of the enriched contact frequencies between centromeric regions and the observed shielding of these regions from chromosome arm interactions.

**Interchromosomal locus–locus contacts**

The interchromosomal contact frequencies in the structure population are correlated with those observed in experiments, with an average Pearson’s correlation of 0.54, which is highly significant ($P < 10^{-15}$) (Fig. 3E,F; Supplemental Fig. 2D). In contrast, the Pearson’s correlation between the random control and experiments is close to nil, and the distinctive contact patterns in the experimental data are completely absent in the random control (Supplemental Fig. 2C).

To examine the effect of limited sampling on the accuracy of chromosomal contact patterns, we compared our initial contact frequency map to maps generated from randomly sampling different proportions of these contacts (Supplementary Material; Supplemental Fig. 3D). In contrast, to intrachromosomal contacts, the correlations between interchromosomal contact patterns are greatly affected by limited sampling. At a sampling rate of 0.1%, we find that the Pearson’s correlation between the two interchromosomal contact maps (even when assuming an ideal model) cannot exceed 0.5. Similar correlation values are observed in the Hi-C experiment when two interchromosomal contact maps are compared that are generated by using two different restriction enzymes (Yaffe and Tanay 2011). In our analysis, the observed correlation value of 0.54 corresponds to a sampling rate of ~0.2%, which is also the order of magnitude that is expected for the experiment (Duan et al. 2010). Thus, the observed correlation coefficient of 0.54 represents a remarkably good agreement between the interchromosomal contact patterns, given that the experimental and computational samplings are finite and cannot be exhaustive.

**Gene localizations**

We now focus on the nuclear locations of individual gene loci. The locations of eight genes have been determined by large-scale fluorescence imaging experiments (Berger et al. 2008). These locations

---

**Figure 2.** Chromosome locations. (A) A sample of 40 chromosome configurations randomly selected from the structure population for the small chromosome 1 (left panel, blue chains), the large chromosome 4 (middle panel, green chains), and the medium-sized chromosome 8 (right panel, gray chains). The chain thickness is reduced to enhance visibility. The chromosomes are depicted in the nucleus with the SPB in pink, the nucleolus in dark blue, and the NE in light blue. (B) Chromosome localization probability densities (LPDs) of chromosomes 1 (left), 4 (middle), and 8 (right panel) plotted along the two principal axes $p$ and $z$ (Supplemental Material). (Lower left) Reference frame for projecting the positions of chromosome points onto the two principal axes, namely, the projection along the central axis $z$ (connecting SPB, nuclear center, and nucleolus), and the radial distance $p$-axis indicating the absolute distance of a point from the central axis. (C) The LPD of chromosome 4 resulting from the ‘‘single chromosome population.’’ The chromosome is subject to all landmark constraints, but structures are generated without the presence of other chromosomes in the nucleus. The density distribution is significantly different from the situation when all chromosomes are present (see B). (D) Excluded volume effect. The difference map between the LPDs of chromosome 4 from the structure population when all chromosomes are present (B, middle) and the single chromosome population as defined in C.

Also these patterns are in excellent agreement with those observed in the conformation capture experiment (Fig. 3F; Duan et al. 2010). The two contact frequency maps are again strongly correlated, with an average row-based Pearson’s coefficient of 0.94 (all $P$-values $<10^{-15}$; Supplemental Material). In contrast, the contact frequency map generated from the random control lacks the cross-shaped patterns (Supplemental Fig. 2C). We now analyze the intra- and interchromosomal locus-pair interaction patterns in more detail.

**Intrachromosomal locus–locus interactions**

Intrachromosomal contact patterns in the structure population and experiments can be divided into three regions (Fig. 4A). Contact frequencies are enriched between regions in the same chromosome arm, as expected for a constrained random polymer chain (blocks c in Fig. 4A). In general, contact frequencies between regions within the same chromosome arm increase with decreasing sequence separation, which is shown by the strong diagonal in the contact frequency maps (Fig. 4A). However, regions located close to the centromere behave very differently. Contacts between subcentromeric regions on opposite sides of the centromere are clearly enriched in frequency, even with increasing chain distance, as can be seen along the line perpendicular to the main diagonal of the contact frequency map (block b in Fig. 4A). Moreover, contact frequencies between subcentromeric regions and regions from the bulk of both chromosome arms are very low (blocks a in Fig. 4A).
are measured with respect to the two principal axes of the nucleus (Methods; Fig. 5A). We determined the two-dimensional (2D) density distributions of the same gene loci in our structure population, allowing for a direct comparison with fluorescence experiments (Berger et al. 2008) (Fig. 5A). The density distribution functions agree well with experiments, in that each locus occupies a well-defined territory. The volumes and shapes of these territories strongly resemble those observed in experiments (Berger et al. 2008). For instance, genes GAL2, HMO1, and SNR17A are located near the nucleolus in the structure population, as seen in the experiment. Interestingly, the structure population places SNR17B (no experiments available) and SNR17A in similar positions near the nucleolus, despite the fact that these genes are located on different chromosomes. Both of these genes are involved in ribosome biogenesis and code the snoRNA U3. Also in agreement with experiments, the distribution patterns of the functionally related genes RPS5 and RPS20 are quite different. For instance, RPS5 positions are significantly more diffuse.

In order to compare quantitatively the relative positions of these eight genes, we measure their median distance along the central axis in the 2D density maps obtained from experimental data and in the structure population. These positions are in excellent agreement (Pearson’s correlation is 0.95, \( P < 10^{-6} \)) (Fig. 5B).

Pairwise telomere distances

It is well known that telomeres are not positioned randomly on the nuclear periphery (Gotta et al. 1996; Bystricky et al. 2005; Berger et al. 2008; Therizols et al. 2010). Fluorescence imaging has revealed that the distance between any two subtelomeres increases gradually with the arm lengths of their chromosomes (Therizols et al. 2010). For a given subtelomere, this relationship is linear. In the structure population, we observe a very similar behavior. More specifically, after applying a change point analysis (Zeileis et al. 2003), we find that the distance between subtelomere pairs as a function of arm length is divided into two linear regimes (Fig. 6). For chromosome arms with lengths up to \( \sim 360 \) kb, the distances observed in our structure population increase with a relatively steep slope. Above \( 360 \) kb, the slope decreases significantly. This behavior is entirely consistent with experiments, and the change in slope has been explained as follows (Therizols et al. 2010). For small arms, the accessible position of a subtelomere at the NE is entirely restricted by the arm length. Hence, the median distance between two subtelomeres increases rapidly with their accessible areas. However, at a certain arm...
length, the subtelomere is able to reach all points on the NE. Further increases in arm length do not dramatically increase the median subtelomere distance.

Interestingly, the change in slope occurs at a slightly different arm length in experiments (~310 kb) (Therizols et al. 2010) compared to our structure population (356–440 kb) (Supplemental Fig. 5a). However, the incompleteness of the experimental data can explain some of this difference. If we only include those chromosomes that are also analyzed in the experiment (Supplemental Fig. 5b), the change in slope in our simulation shifts to 309–327 kb in remarkably good agreement with the experiment.

**Telomere clusters**

To identify subtelomere clusters in the structure population, for each subtelomere we calculated the fraction of structures with at least one other subtelomere within 250 nm. In agreement with another experiment (Therizols et al. 2010), such small subtelomere distances are infrequent (1%–3%) for most chromosomes (Supplemental Fig. 6A). However, a few co-locations are observed more frequently between relatively short chromosome arms, namely, 1R:1L, 6R:6L, 1R:9R, and 3L:3R. These pairings occur in 12%–20% of the population, also in agreement with experiments (Bystricky et al. 2005; Schober et al. 2008; Therizols et al. 2010). For example, the pairs 3R:3L and 6R:6L were recently reported to form significant but transient interactions, leading to the formation of chromosome loops (Bystricky et al. 2005; Schober et al. 2008; Therizols et al. 2010). In general, we find that as the length difference between chromosome arms grows more pronounced, the probability of their telomeres being co-localized decreases. Interestingly, the volume exclusion effect has a pronounced effect on the co-location frequency of telomeres. For small and also large chromosomes, the volume exclusion effect increases significantly the co-location frequency, while for medium-sized chromosomes the opposite is observed by decreasing the co-location frequency (Supplemental Fig. 6B). For instance, the fraction of co-located telomeres increases by almost 20% for the small chromosome 6 upon the presence of all other chromosomes in the nucleus, while it decreases by 60% for the medium-sized chromosome 8.

**Co-localization of functionally related loci**

Next, we investigate whether functionally related gene loci are co-localized in the structure population. First, we compare the 3D spatial distributions of early and late replication start sites in the structure population. These sites are distributed across all chromosomes (Fig. 7, right panels). Experimental evidence exists that early replication sites are spatially clustered during interphase (Di Rienzi et al. 2009; Duan et al. 2010).

In each structure of the population, we calculate the mean pairwise distance between all early replication sites. The frequency distribution derived from these mean pairwise distances is compared to a distribution chosen from randomly selected sites in the genome. We observe significant spatial clustering of the early replication sites (Fig. 7A), in the sense that their mean pairwise distances are significantly less than would be expected from randomly selected sites (Stouffer’s Z-transform [Stouffer et al. 1949] tests z-scores <$-160$; Supplemental Material). This observation holds for all three sets of early replication origins identified in the literature (Feng et al. 2006; McCune et al. 2008; Sekedat et al. 2010). Remarkably, for late replication sites we see the opposite effect: a statistically significant increase in the mean pairwise distances between late replication sites compared with the background. It appears that, on average, early replication start sites are closer to the centromere on the chromosome sequence compared with the late start sites (all P-values <10^{-3} for the three data sets) (Supplemental Material).

We also analyzed the spatial positions of all tRNA gene loci in the genome, which have been observed to cluster in experiments (Thompson et al. 2003; Duan et al. 2010). Again, we observe a statistically significant decrease in the pairwise distances between tRNA loci (Fig. 7B) compared with randomly picked loci.

Our observations clearly indicate that the chromosomal locations of these specific sites are not randomly distributed over the genome; they are positioned in such a way that early replication sites have a higher probability of being co-localized when the chromosome chains behave as random polymer chains that are subject to nuclear landmark constraints.

**Discussion**

In this paper, we demonstrate that purely random configurations of tethered chromosomes reproduce in a statistical manner all the experimental hallmarks of genome organization in Saccharomyces cerevisiae. Specifically, random configurations generate structural features that agree remarkably well with (1) the highly specific...
interaction patterns between individual chromosomes, chromosome regions, and chromosome folding patterns obtained by genome-wide conformation capture experiments (Duan et al. 2010); (2) the emergence, shape, and position of individual gene territories revealed by probability maps from fluorescence experiments (Berger et al. 2008); (3) the distribution of median distances between telomeres; (4) the relative frequencies of telomere co-localizations observed in imaging experiments (Brystricky et al. 2005; Schober et al. 2008; Therizols et al. 2010); and even (5) the physical proximity of functionally related gene loci, including early replication sites and tRNA gene loci.

In addition to chromosome tethering, the main organizing factor is a volume exclusion effect, as a result of the competition of all the chromosomes for the limited nuclear space. The fact that the chromosome arms have different lengths gives rise to important nuances of organization and implies that the locations of a gene or chromosome territory depends on all the other chromosomes. Therefore, the gene territory position and specific interaction patterns of a given gene locus is determined not only by its chromosome sequence position and the arm lengths of its own chromosome but also by the total number and the relative arm lengths of all other chromosomes. The volume exclusion effect can even create counterintuitive effects. For instance, for small and large chromosomal volumes the exclusion effect leads to an increase in the frequency with which subtelomeres on the same chromosome are in proximity to each other, while for medium-sized chromosomes a decrease is observed.

Our findings have several important consequences. First, we show that a small number of purely geometrical constraints on otherwise randomly configured chromosomes can lead to a highly structured 3D genome organization. Second, the hallmarks of genome organization can be explained without calling on specific molecular interactions between chromatin regions or chromatin-bound proteins. For instance, random chromosome encounters can also statistically explain the spatial features often attributed to an apparent Rabl-like chromosome folding, which refers to the back-folding of subcentromeric chromosome regions so that chromosome arms appear juxtaposed. This pattern is mainly caused by the volume exclusion effect (Fig. 4). In response to the competition for the limited space around the SPB, chromosome regions on both sides of the centromere show a statistical preference for bending toward each other. When averaged over the entire cell population, this tendency gives rise to the distinctive cross-shaped intrachromosomal contact patterns observed in experiments and in our structure population. However, most individual structures will not exhibit simultaneously all the features of such an apparent Rabl-like fold. We therefore emphasize that the data should be explained as a statistical preference for chromosome contacts but not necessarily be interpreted as a stable chromosome fold. An interesting prediction of this model is that the Rabl-like subcentromeric contact pattern should not be expected in yeast species if the number of chromosomes was considerably smaller even if the chromosomes were all tethered to nuclear landmarks. Although S. pombe and S. cerevisiae have similar genome sizes, the former has only three chromosomes. The prediction is sustained: In genome-wide conformation capture experiments, Schizosaccharomyces pombe yeast does not show the cross-shaped intrachromosomal contact patterns characteristic of this type of folding (Tanizawa et al. 2010).

Another remarkable result is that the early replication sites in our structure population have a high probability of being in close proximity compared with the background distribution of pairwise separations. In contrast, late replication sites have a lower probability of being colocated compared with randomly selected sites. This difference may help regulate a naturally occurring order on replication timing. The existence of these and other co-location patterns may indicate that the relative positions of affected loci in the chromosome were selected by evolution. Due to excluded volume effects, the spatial position of a gene in the nucleus is not only modulated by its relative sequence position in its own chromosome, but also by the relative arm lengths and the total number of all other chromosomes in the nucleus.

We also note that our study provides additional evidence for the existence of a chromatin fiber in the yeast interphase nucleus with length and density properties similar to the 30-nm fiber. We created an alternative structure population consistent with a 10-nm chromatin fiber, and the statistical results do not agree with the described experimental evidence.

Finally, we believe that our results point toward a considerable structural variability of genome structures among individual cells. Each structure in our population not only differs considerably from the “average conformation” but also from all the other structures in the population (<0.3% of loci contacts are shared between any two structures; Supplemental Material). No single-genome structure or population-averaged structure is representative of the population. Although the true structural variability is unknown, our results indicate that a single structural model cannot adequately reflect all the spatial features of the genome. It is crucial to analyze genome structures from a statistical rather than an individual standpoint. Structural analysis should be performed by generating a population of 3D genomes, which represent the spectrum of all possible chromosome configurations consistent with the data. The structural organization of the genome can then be interpreted sta-
of our analysis. The compaction ratio will slightly change the total number of beads in a model. The restraints are expressed as pseudo potential energy terms derived from experimental information. To optimize the structure, the scoring function is minimized to a score of zero. The scoring function is written as

\[
S(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_{i=1}^{N-1} \mathcal{U}_{\alpha}(\mathbf{r}_i, \mathbf{r}_{i+1}) + \sum_{i=1}^{N} \mathcal{U}_{\beta}(\mathbf{r}_i) + \sum_{\mathbf{g} \in \mathbb{Z}^3} \mathcal{U}_{\gamma}(\mathbf{r}_{\mathbf{g}}) + \sum_{\mathbf{i}, \mathbf{j} \in \mathbb{Z}^3} \mathcal{U}_{\delta}(\mathbf{r}_i, \mathbf{r}_j) + \mathcal{U}_{\text{cen}}(\mathbf{r}) + \mathcal{U}_{\text{nuc}}(\mathbf{r}) + \mathcal{U}_{\text{exc}}(\mathbf{r}) + \mathcal{U}_{\text{nucl}}(\mathbf{r}) = 0,
\]

where \( \mathbf{r}_i \in \mathbb{R}^3 \) is the coordinate vector of bead \( i \), and \( N \) is the total number of beads in a model. The restraints are expressed as pseudo potential energy terms \( \alpha, \beta, \gamma, \) and \( \delta \) are subsets of specific beads in the chromosome chains that share certain properties. More specifically, \( \alpha \) is the set of beads assigned to the last bead of every chain, \( \beta \) is the set of beads assigned to centromeres, \( \gamma \) is the set of beads assigned to telomeres, and \( \delta \) is the set of all beads flanking rDNA repeat regions.

### Chromatin chain restraint \( \mathcal{U}_{\alpha} \)

Two consecutive beads in a chromatin chain are restrained to be at a distance of 30 nm (Table 1).

### Chromatin chain excluded volume restraint \( \mathcal{U}_{\text{exc}} \)

Overlap between beads is prevented by imposing excluded volume restraints for all bead pairs (Table 1).

### NE restraint \( \mathcal{U}_{\text{nuc}} \)

All chromatin beads must remain within the nucleus, defined as a sphere with radius \( R_{\text{nuc}} = 1 \) micron (Table 1).

### Centromere localization restraint \( \mathcal{U}_{\text{cen}} \)

All the centromeres cluster near the SPB, which is the microtubule organization center in the yeast nucleus (Jin et al. 2000). The centromeric regions are attached to the SPB through microtubules up to 300 nm in length (O'Toole et al. 1999). Accordingly, all beads representing centromeric regions are restricted to a spherical volume with a radius 300 nm, centered on the SPB (Fig. 1). We follow ex-
experimental evidence from fluorescence imaging and place the centromere localization volume on the central axis, close to the NE (Fig. 1, scheme in top right panel; Berger et al. 2008; Therizols et al. 2010). Beads representing telomerese are positioned in the vicinity of the NE (Table 1; Fig. 1, thin gray shell of 50-nm thickness).

**Telomere localization restraint $U_{tel}$**

Telomerese have a high probability to be located near the nuclear periphery (Berger et al. 2008; Therizols et al. 2010). Beads representing telomerese are positioned in the vicinity of the NE (Table 1; Fig. 1, thin gray shell of 50-nm thickness).

**Nucleolus localization restraint $U_{nucl}$**

The rDNA is located on chromosome 12 and consists of 150–200 tandem repeats of 9.1 kb length each (Kim et al. 2006; Taddei et al. 2010). All rDNA regions are found in the nucleolus. Because no conformation capture data or fluorescence imaging data are available for the rDNA genes, we do not explicitly resolve the chromatin fiber within the nucleolus. Instead, we anchor the two beads at the beginning and end of the rDNA repeat region (i.e., positions 458 kb from the left telomere and 620 kb from the right telomere in the sequence of chromosome 12) to the surface of the nucleolus (Table 1; Fig. 1).

**Chromatin persistence length**

During the optimization process, we imposed an angular restraint between each set of three consecutive beads to reproduce the desired chain stiffness. The constraint is expressed as a harmonic potential:

$$U_{\text{angle}} = \frac{1}{2} k_{\text{angle}} \sum_{i=1}^{N-2} \angle i \times (1 - \frac{r_{i+1} - r_{i} - r_{i-1} - r_{i+2}}{r_{i+1} - r_{i} - r_{i-1} - r_{i+2}})^2,$$

for $i, i + 1$, and $i + 2$ on the same chain.

This restraint is considered only when calculating gradient forces during the optimization process. It makes no contribution to the total score of a model (below). With a force constant of $k_{\text{angle}} = 0.2$ kcal/mol, we obtain chromatin chains that behave like random polymers with a persistence length between 47 and 72 nm (the average is $61.7 \pm 7.7$ nm) (Supplemental Fig. 7), consistent with experiments. Estimated values for the persistence length from experiments fall between 30 and 220 nm (Cui and Bustamante 2000; Bystricky et al. 2004; Langowski 2006).

**Optimization**

The optimization is performed using a combination of simulated annealing molecular dynamics and the conjugate gradient methods implemented in the Integrated Modeling Platform (IMP; http://www.integrativemodeling.org) (Alber et al. 2007a,b, 2008; Russel et al. 2012). An individual optimization starts with an entirely random bead configuration, followed by an initial optimization of the structure. Next, we apply simulated annealing protocols to entirely equilibrate the genome configuration. Finally, conjugate gradient optimization ensures that all constraints are satisfied, leading to a structure with score zero. Many independent optimizations are carried out to generate a population of 200,000 genome structures with a total score of zero, hence consistent with all input data. A comparison between the frequency maps of two independently calculated populations, each with 100,000 structures, showed that our genome structure population is highly reproducible (Pearson’s correlation between the contact frequency maps of the two populations is 0.999).
Control population
We also generated a control population of 25,000 structures without chromosome tethering constraints and nucleolus excluded volume constraints. Otherwise, the chromosomes are constrained in a manner identical to the full simulation.

Analysis
The analysis of the structure population and all statistical tests are described in great detail in the Supplemental Material.

Acknowledgments
We thank Dr. Oscar Aparicio, Dr. Susan Forsburg, Dr. M.S. Madhusudhan, Dr. Reza Kalhor, Zachary Frazier, Dr. Jasmine Zhou, Madhusudhan, Dr. Reza Kalhor, Zachary Frazier, Dr. Jasmine Zhou, Dr. Wenyuan Li, and Chao Dai for useful discussions and comments on the manuscript, and the IMP developers for supporting the IMP modules. This work is supported by the Human Frontier Science Program, grant RGY0079/2009-C to F.A.; the Alfred P. Sloan Research Foundation grant to F.A.; NSF CAREER grant 1150287 to F.A.; NIH grants HL076334, GM064624, and GM077320 to L.C.; NIH grant 1R01GM096089 to F.A.; and NIH grant 2U54RR022220 to F.A. and L.C. F.A. is a Pew Scholar in Biomedical Sciences, supported by the Pew Charitable Trusts.

References


chromosome replication: genomewide replication in clb5Δ <i>Saccharomyces cerevisiae</i>. <i>Genetics</i> 180: 1833–1847.

Received July 22, 2011; accepted in revised form April 12, 2012.