Computational Predictions of Structures of Multichromosomes of Budding Yeast

(Accepted, Conf Proc IEEE Eng Med Biol Soc. 2014)

Gamze Gürsoy[1], Yun Xu[1], and Jie Liang[1]

Abstract— **Knowledge of the global architecture of the cell nucleus and the spatial organization of genome is critical for understanding gene expression and nuclear function. Single-cell imaging techniques provide a wealth of information on the spatial organization of chromosomes. Computational tools for modelling chromosome structure have broad implications in studying the effect of cell nucleus on higher-order genome organization. Here we describe a multichromosome constrained self-avoiding chromatin model for studying ensembles of genome structural models of budding yeast nucleus. We successfully generated a large number of model genomes of yeast with appropriate chromatin fiber diameter, persistence length, and excluded volume under spatial confinement. By incorporating details of the constraints from single-cell imaging studies, our method can model the budding yeast genome realistically. The model developed here provides a general computational framework for studying the overall architecture of budding yeast genome.**

I. INTRODUCTION

Nucleus is the control center of the cell. Properties and organization of the nucleus are important determinants of the behavior of a cell. Understanding how chromosomes are organized in the cell nucleus and the effect of nuclear structures on genome organization are essential for gaining insights into the mechanism of gene activities, nuclear functions, and maintenance of cellular epigenetic state [1].

Single-cell imaging techniques provide a wealth of information about the organization of genome and nuclear structures in budding yeast nucleus. The spatial clustering of certain genomic elements is a key property of budding yeast nucleus. For example, telomeres are anchored to the nuclear envelope through protein factors [2], [3], [4], [5], and centromeres are clustered adjacent to the spindle pole body (SPB) [6]. Morever, ribosomal DNA (rDNA) repeats are located in the sub-nuclear compartment called nucleolus [7], [8], [9], where rDNA transcription takes place [7], [9], [10], [5], [11].

*This work was supported by NIH Grants GM079804, GM086145 and NSF Grants DBI 1062328 and DMS-0800257. This work was also funded by the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust

[1]G. Gürsoy is with the Bioinformatics Program, Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, 60607, USA ggurso2@uic.edu

[1]Y. Xu is with the Bioinformatics Program, Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, 60607, USA yxu7@uic.edu

[1]J. Liang is on Faculty of Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, 60607, USA jliang@uic.edu

Computational modelling of chromosomes has become increasingly important for understanding spatial organization of cell nucleus and mechanism of gene regulation. A number of computational methods have been developed to study the yeast genome [12], [13]. A useful approach is to recapitulate important folding features of yeast genome using only a minimum number of experimental constraints from single-cell imaging data [13]. In a previous study, the problem is formulated as that of an optimization of finding a 3D genome model satisfying all of the given minimum constraints [13]. Solution was obtained using a psuedo energy function [13]. This approach was very successful in reproducing some of the known experimental facts of yeast genome based on population properties of model chains [13]. However, deriving an optimal solution that satisfies all the constraints is extremely difficult.

Here we describe a novel model of multichromosome constrained self-avoiding chromatin by sequentially growing independent ensembles of multiple interacting chromatin chains that satisfy the experimental constraints, with the goal to recapitulate the genome organization in budding yeast genome and to predict important interactions that have not been captured experimentally. Our results successfully reproduced a number of hallmarks of the genome organization in budding yeast nucleus, such as telomere and centromere co-localization in details.

II. MODELS AND METHODS

A. Polymer model and physical properties

The architecture of budding yeast nucleus consists of the spindle pole body, the nuclear envelope, the nucleolus, and 16 chromosomes [13]. The nucleus diameter is $\sim$2 μm [16], [10], [17], [15]. Imaging experiments showed that the SPB and nucleolus are located at opposite ends of the nucleus (Fig. 1) [10].

In our model, a chromatin fiber is modeled as a self-avoiding polymer chain consisting of beads in accordance with previous studies [18]. Each bead has a diameter of 30 nm [19], [20], and corresponds to 3 kb of DNA [21], [22]. Every 5 beads form a persistent unit, and correspond to a persistence length L_p of $\sim$ 150 nm [22]. The entire budding yeast genome is represented by a total of $800L_p$, which are divided into 16 different chromosomes.

B. Chain growth by Geometrical Sequential Importance Sampling (GSIS)

The chromosome chains of the yeast genome in three-dimensional space are generated by a chain growth approach [23], [24]. The location of centeromeres and telomeres are mapped onto polymer chains according to their genomic locations. Each chromosome is divided into two chains from their centromeres as left arm and right arm. If a chromosomal arm chain consists of n persistence units, with the location of the i-th persistence unit denoted as $x_i = (a_i, b_i, c_i) \in \mathbb{R}^3$, the configuration x of a full chromosomal arm chain with n persistence units is:

$$x = (x_1, \cdots, x_n).$$

To generate a chromosomal arm, we grow the chain one persistence unit at a time, ensuring the self avoiding property along the way, namely, $x_i \neq x_j$ for all $i \neq j$. We use a $k = 640$-state off-lattice discrete model (see [23], [24], [25] for more details). The new persistence unit added to a partial chain is placed at x_{t+1}, which is a persistence length L_p distance away from current persistence unit located at x_t. x_{t+1} is taken from one of the unoccupied k-sites neighboring x_t with a probability of growth $g(x)$ according to their distance to the target experimental constraints. This selection introduce bias for sampling from the target distribution $\pi(x)$, we therefore keep track of the bias and assign each successfully generated genome a proper weight $w(x) = \pi(x)/g(x)$. Details can be found in references [23], [24], [25]. During the chain generation process, we randomly choose a chromosomal arm, place its corresponding centromere to a random location in the SPB and grow the chain towards the target constraint. We repeat this process until all 32 chromosomal arms are completely generated.

C. Target distribution and geometrical constraints

The first persistence unit of each chromosomal arms (centromeres) are randomly attached to any location in the SPB. Each partial chromosomal arm x_t is grown from centromeres according to the target distribution $\pi(x_t)$ based on the geometrical constraints derived from experimental data. During the growth process, (i) violation of self-avoiding property, (ii) violation of non-rDNA elements being located in nucleolus, and (ii) violation of any persistence units being located outside of nucleus are not allowed. The target distribution $\pi(x)$ is the Boltzmann distribution of all chromosome chains that satisfy the experimental constraints within the given confinement. The target distribution $\pi(x_t)$ of a partial chain follows Boltzmann distribution as

$$\pi(x_t) = \exp(-\alpha u(x_t)/k_B T),$$

$$u(x_t) = d^2_{\mathrm{t-target}},$$

$d_{\mathrm{t-target}}$ is the distance between the persistence unit t and the nuclear envelope if t is a telemore. $d_{\mathrm{t-target}}$ is the distance between the persistence unit t and the nucleous if t contains rDNA repeat. α is a constant and k_B is the Bolztmann factor.

D. Growth function

Since it is impossible to sample multiple chains that satisfy the constraints with a random growth scheme, biasing the chain growth at each persistence unit that is not constrained with a designed probability of selecting an unoccupied site according to our growth function. For a persistence unit t, which is not the telomere or an rDNA repeat, the probability of growth from the position of x_{t-1} to any unoccupied site k is

$$g(x_{t_k}) = \exp(-\beta u'(x_{t_k})/k_B T),$$

$u'(x_{t_k}) = d^2_{\mathrm{t_k-target}}$ if $d_{\mathrm{t_k-target}} \geq m \times L_p$ and $u'(x_{t_k}) = 0$ otherwise. Here, $d_{\mathrm{t_k-target}}$ is the distance between the candidate position of the t_kth persistence unit to a randomly predetermined position on nuclear envelope if the target persistence unit is a telemore. $d_{\mathrm{t_k-target}}$ is the distance between the candidate position of the t_kth persistence unit to a randomly pre-determined position on nucleolus if the target persistence unit is an rDNA repeat. m is the number of persistence units between tth persistence unit and the target. Our algorithm is implemented in C++.

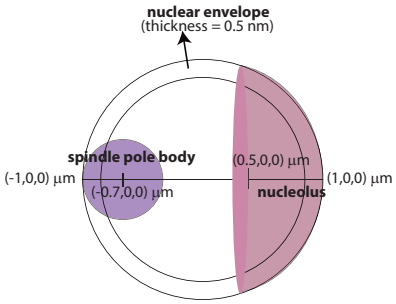

Fig. 1. Schematic representation of budding yeast nuclear architecture

III. Results and Discussion

A. Analysis of structural properties of budding yeast genome

We generated an ensemble of ~200,000 genome structural models and analyzed the average structural properties. The spatial properties of model genomes are compared to the available experimental data. First, we carried out a violation analysis by assessing (i) centromere localization and clustering in the SPB as the experiments suggested [6], and (ii) inaccessibility of nucleolus to chromosomes, except the regions with rDNA repeats [7], [8], [9]. We calculated the spatial positions of all centromeres and the chromosomal units with and without rDNA repeats. Results showed that all the centromeres are indeed clustered in the SPB and only persistence units with rDNA elements are located in the nucleolus (Fig. 2). These results verify that our model conform with experimental constraints, which were set as important restraints in the beginning of our chain growth procedure.

B. Validation of model predictions

We calculated the frequency of pairwise interactions in our model ensemble and in chromosome conformation capture

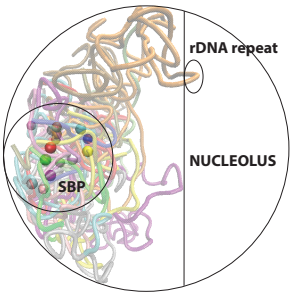

Fig. 2. Centromere co-localization and clustering in the spindle pole body and rDNA repeat localization in the nucleolus are captured. Centromeres are shown as solid spheres.

experiments [12] at 15 kb resolution. The Pearson correlation coefficient is 0.92 between the predicted frequency matrix and the experimentally measured frequency matrix (Fig. 3). This suggests that physical architecture of yeast genome largely arise from topological constraints of cell nucleus and excluded volume effect. These results are also consistent with our previous work, which suggests that folding landscape of human genome is dictated by confinement of cell nucleus [25]. The problem of recovering the spatial organization of yeast genome with a minimum number of constraints has been studied by Tjong et. al [13], and a Pearson correlation of 0.94 is obtained at 32 kb resolution. Our correlation is comparable to that of Tjong et. al, but at a higher resolution of 15 kb. In addition, the self-avoiding property of chromatin fiber is enforced in our model, while self-crossing is allowed in Tjong et. al [13]. This result shows that we can recover the overall organization of yeast genome at a higher resolution with a realistic physical model.

We further tested predictions of our model with the available experimental data. We examined (i) inter- *vs.* intra-chromosomal contacts of centromeres, and (ii) telomere clusters. According to chromosome conformation capture experiments [12], centromeric regions have higher propensity to engage in inter-chromosomal interactions than in intra-chromosomal interactions. To test the predictive power of our method, we examined whether a similar pattern is seen in our model genome. We calculated the pairwise distances between persistence units. Our results show that there is a clear pattern in the genome demonstrating that centromeres do not interact with their chromosomal arms (Fig. 4).

To quantify this finding, we calculated the number of intra- and inter-chromosomal contacts that centromeric persistence units engage in by setting the interaction threshold to 200 nm. Our results show that almost all the centromeres engage in more inter-chromosomal interactions than intra-chromosomal interactions (Fig. 4).

To identify telomere clustering in the model genome, we divided the nuclear envelope into 3D grids as each grid occupies a $0.5 \times 0.5 \times 0.5$ μm^3 volume. We plotted the locations of each telomere in the nuclear envelope (Fig. 5). We found that there are sub-clusters of telomeres, which are especially densely located in grid 1 and 2 in Fig. 5. Experiments also showed that inter-chromosomal telemore

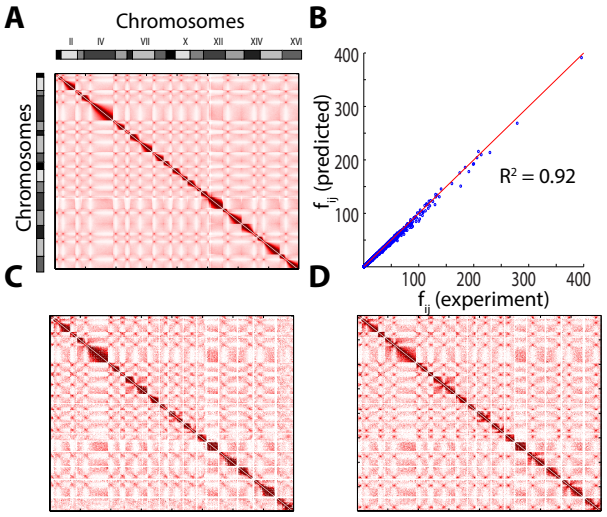

Fig. 3. (A) Heatmap of frequency of pairwise interactions in the model ensemble. All persistence units of the chains are taken into consideration. (B) Correlation coefficient between frequency of pairwise interactions in model ensemble and in experiments. Persistence unit pairs that have experimental frequencies are included. (C) Heatmap of frequency of pairwise interactions in the model ensemble, where persistence unit pairs that have experimental frequencies are included. (D) Heatmap of frequency of pairwise interactions in chromosome conformation experiments [12].

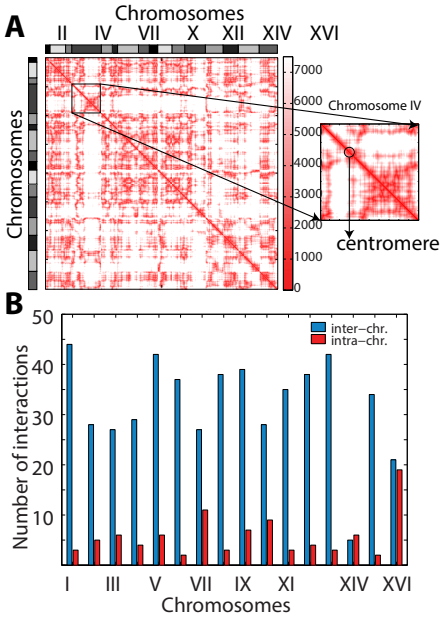

Fig. 4. (A) Heatmap of pairwise distances between persistence units for the entire genome and for Chr IV. The distances are color coded. As the distance between persistence units decreases, the pairwise distance is shown as darker red, indicating a physical interaction between those units. (B) Number of intra- and inter-chromosomal contacts of centromeric persistence units per chromosome. All chromosomes except Chr 14 engage in higher inter-chromosomal interactions at their centromeres.

co-locallization is observed between Chromosome 1 and 9 [26]. Strikingly, we found that telomere of Chr 1 and the telemores of Chr 9 are clustered in the same grid, namely within 0.5 μm of each other. Furthermore, we predicted more intra- and inter-chromosomal telomere subclustering (Fig. 5). Additional experimental investigations will be fruitful to validate these findings.

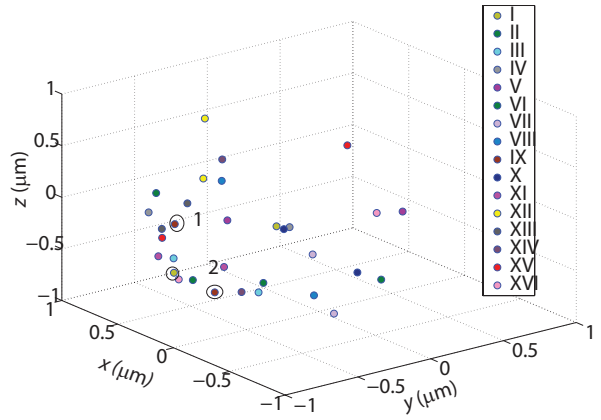

Fig. 5. Telomeric persistence units are located on the nuclear envelope and forms sub-clusters. Nuclear envelope is divided into grids and circled telomeres belong to Chr 1 and 9.

In summary, predictions on pairwise interaction frequencies, centromere interactions, and telomere clustering are all consistent between our model and known experimental facts. This is achieved without requiring additional biochemical data. Our results show that a few spatial constraints that come from internal architecture of yeast nucleus is sufficient to enable the construction of a three-dimensional structural model of budding yeast genome. Our results suggest that the spatial organization of budding yeast genome is dictated by spatial confinement of cell nucleus and physical localization of nuclear structures such as SPB and nucleolus as suggested by previous studies [13].

IV. Conclusion

Here we present a novel method to realistically model three-dimensional structure of budding yeast genome, which is based on a geometrically constrained chain growth method. These geometrical constraints are derived from the physical architecture of cell nucleus and relative positioning of centromeres, telomeres and rDNA repeats. Our method allows generation of ensemble of model genomes, and results in spatial structural properties of yeast genome are consistent with experimental data.

V. Acknowledgments

Support from NIH Grants GM079804, GM086145 and NSF Grants DBI 1062328 and DMS-0800257 and the Chicago Biomedical Consortium are gratefully acknowledged.

References

[1] Fraser P and Bickmore W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447(7143):413–417.

[2] Hediger F, Neumann F R, Van Houwe G, Dubrana K, and Gasser S M (2002) Live imaging of telomeres: yKu and sir proteins define redundant telomere-anchoring pathways in yeast. *Current Biology*, **12**(24):2076–2089.

[3] Taddei A, Hediger F, Neumann F R, and Gasser SM (2004) The function of nuclear architecture: a genetic approach. *Annual Review of Genetics*, **38**:305–345.

[4] Taddei A, Van Houwe G, Nagai S, Erb I, van Nimwegen E, and Gasser S M (2009) The functional importance of telomere clustering: global changes in gene expression result from SIR factor dispersion. *Genome Research*, **19**(4):611–625.

[5] Mekhail K and Moazed D (2010) The nuclear envelope in genome organization, expression and stability. *Nature Reviews*, **11**(5):317–328.

[6] O'Toole E T, Winey M and McIntosh J R (1999) High-voltage electron tomography of spindle pole bodies and early mitotic spindles in the yeast Saccharomyces cerevisiae. *Molecular Biology of the Cell*, **10**(6):1059–1524.

[7] Yang C H, Lambie E J, Hardin J, Craft J, and Snyder M (1989) Higher order structure is present in the yeast nucleus: autoantibody probes demonstrate that the nucleolus lies opposite the spindle pole body. *Chromosoma*, **98**(2):123–128.

[8] Dvorkin N, Clark M W, and Hamkalo B A (1991) Ultrastructural localization of nucleic acid sequences in saccharomyces cerevisiae nucleoli. *Chromosoma*, **100**(8):519–523.

[9] Bystricky K, Laroche T, van Houwe G, Blaszczyk M, and Gasser SM (2005) Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *The Journal of Cell Biology*, **168**(3):375–387.

[10] Berger A B, Cabal G G, Fabre E, Duong T, Buc H, Nehrbass U, Olivo-Marin J-C, Gadal O, and Zimmer C (2008) High-resolution statistical mapping reveals gene territories in live yeast. *Nature Methods*, **5**(12):1031–1037.

[11] Taddei A, Schober H, and Gasser S M (2010) The budding yeast nucleus. *Cold Spring Harbor Perspectives in Biology*, **2**(8):a000612.

[12] Duan Z, Andronescu M, Schutz K, McIlwain S, Kim Y J, Lee C, Shendure J, Fields S, Blau C A, and Noble W S. (2010) A three-dimensional model of the yeast genome. *Nature*, **465**(7296):363–367.

[13] Tjong H, Gong K, Chen L, and Alber F (2012) Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Research*, **22**(7):1295–1305.

[14] Ba D, Sanyal A, Lajoie B R, Capriotti E, Byron M, Lawrence J B, Dekker J, and Marti-Renom M A (2011) The three-dimensional folding of the a-globin gene domain reveals formation of chromatin globules. *Nature Structural & Molecular Biology* **18**(1):107–114.

[15] Gasser S M (2002) Visualizing chromatin dynamics in interphase nuclei. *Science* **296**:1412–1416.

[16] Chubb J R and Bickmore W A (2003) Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell* **112**:403–406.

[17] Meister P, Gehlen L R, Varela E, Kalck V and Gasser S M (2010) Visualizing yeast chromosomes and nuclear architecture. *Methods in Enzymology* **470**:535–567.

[18] Barbieri M, Chotalia M, Fraser J, Lavitas L M, Dostie J, Pombo A, and Nicodemi M (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *PNAS*, **109**(40):16173–16178.

[19] Finch J T and Klug A (1976) Solenoidal model for superstructure in chromatin. *PNAS* **73**(6):1897–1901.

[20] Bystricky K, Heun P, Gehlen L, Langowski J, Gasser S M (2004) Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *PNAS* **101**:16495–16500.

[21] Gerchman S E, Ramakrishnan, V (1987) Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy. *PNAS* **84**:7802–7806.

[22] Wedemann G, Langowski J (2002) Computer simulation of the 30-nanometer chromatin fiber. *Biophysical Journal* **82**:2847–2859.

[23] Liang J, Zhang J, Chen R (2002) Statistical geometry of packing defects of lattice chain polymer from enumeration and sequential Monte Carlo method. *The Journal of Chemical Physics* 117:3511–3521.

[24] Zhang J, Chen R, Tang C, Liang J (2003) Origin of scaling behavior of protein packing density: A sequential monte carlo study of compact long chain polymers. *The Journal of Chemical Physics* **118**:6102.

[25] Gürsoy G, Xu Y, Kenter A L, Liang J (2014) Spatial confinement is a major determinant of folding landscape of human genome. *Nucleic Acids Research* **in press**.

[26] Therizols P, Duong T, Dujon B, Zimmer C, Fabre E (2010) Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *PNAS* **107**:2025–2030.