

REVIEW

Computational inference of physical spatial organization of eukaryotic genomes

Bingxiang Xu and Zhihua Zhang*

CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

* Correspondence: zhangzhihua@big.ac.cn

Received July 27, 2016; Accepted August 24, 2016

Background: Chromosomes are packed in the cell's nucleus, and chromosomal conformation is critical to nearly all intranuclear biological reactions, including gene transcription and DNA replication. Nevertheless, chromosomal conformation is largely a mystery in terms of its formation and the regulatory machinery that accesses it.

Results: Thanks to recent technological developments, we can now probe chromatin interaction in substantial detail, boosting research interest in modeling genome spatial organization. Here, we review the current computational models that simulate chromosome dynamics, and explain the physical and topological properties of chromosomal conformation, as inferred from these newly generated data.

Conclusion: Novel models shall be developed to address questions beyond averaged structure in the near future.

Keywords: 3D genome; models; simulation

INTRODUCTION

The compact conformation of nuclear genome is critical to nearly all intranuclear cellular processes, including transcription, DNA replication and DNA damage repair [1,2]. Currently experimental technologies, such as fluorescence *in situ* hybridization (FISH) and chromosome conformation capture (3C) have improve our understanding of genome spatial organization [3–5]. The genome can be roughly organized into hierarchical structures in multiple scales. At the bottom, genomic DNA wraps around nucleosomes in approximately 147 base pairs, forming a 10 nm basic unit. These 10 nm particles are, in turn, connected by linker DNA, further forming a secondary structure known as nucleosome fiber, which has been observed by cryogenic electron microscopy (cryo-EM) as a 30 nm double helix structure [6], with the possibility of other formations existing *in vivo* [7]. Nucleosome fiber can further fold into a tertiary genomic structure, but this is much less characterized by current technologies. Several models have been proposed based on high-throughput 3C data, e.g., the “bottle-brush” model [8,9]. Above the tertiary structure, the genome has been observed to occupy mutually exclusive nuclear

space for each individual chromosome [10], even between sister chromosomes [11]. Both primary and secondary genome structure can be observed by current microscopy techniques, but technologies to directly observe details beyond the secondary genome structure are not available. Consequently, computation inference of chromatin structure has become the key toward understanding the principles of genome spatial organization. In this review, we introduce computational models of chromatin structure with indirect structural measurement. We first introduce basic polymer models with scales from several kilobases to chromosome-scale, followed by introducing current efforts to model structural ensembles in cell populations. Finally, we discuss integrative models on multi-omics data for 3D genome structures.

It is natural to model chromatin as linear polymers showing how DNA is packed into an array of nucleosomes in eukaryotic cells, notwithstanding variability in the size of monomers from tens to millions of base pairs. Let a chromosome consist of N straight and elastic segments; then the positions of segment ends are denoted by vectors $\{r_i\}_{i=0}^N$, and the segment i can be written as vector $s_i = r_i - r_{i-1}$, $i = 1, 2, \dots, N$. Because different factors control the conformation between small and large

scales, we review the models in the two separate groups.

MOLECULAR DYNAMIC MODELS FOR LOCI AND SMALL CHROMOSOMES

At small scale where the size of monomers varies from several bases to several kilobases, chromatin conformation is largely affected by intramolecular stretching, bending, and twisting, as well as hydrophilic and electrostatic exclusions [12–14]. We describe the models beginning with the definition of energy functions. The stretching energy, $E_i^{(s)}$, of any given segment s_i is defined by

$$\frac{E_i^{(s)}}{k_B T} = \frac{1}{2(l_0 \delta)^2} (l_0 - \|s_i\|)^2, \quad (1)$$

where k_B is the Boltzmann constant, T is the absolute temperature, l_0 is the equilibrium length of the segment, and δ is the stiffness parameter [12]. The bending energy,

E_i^b , for adjacent segments s_{i-1} and s_i , is defined by

$$\frac{E_i^b}{k_B T} = \frac{1}{2} k_\theta \langle s_i, s_{i+1} \rangle^2, \quad i = 1, 2, \dots, N-1, \quad (2)$$

where k_θ is the bending elasticity of the chain and $\langle *, * \rangle$ denotes the torsion angle between the two vectors [15].

The twisting energy, $E_i^{(t)}$, for segments $s_i, i = 2, 3, \dots, N-1$ is defined by

$$\frac{E_i^{(t)}}{k_B T} = \frac{C}{2k_B T l_0} \tau_i^2, \quad (3)$$

where C denotes the torsional rigidity of the segment, and τ_i denotes the twist angle which is given as the angle between the projections of s_{i-1} and s_{i+1} onto the plane vertical to s_i [12]. The exclusion energy by hydrophilic and electrostatic effect is described in detail in [12] and can be simplified by the repulsive part of the Lennard–Jones potential as

$$\frac{E_{ij}^{(e)}}{k_B T} = \begin{cases} 4I \left(2^{1/6} - \frac{d_{ij}}{\sigma} > 0 \right) \left(\left(\frac{\sigma}{d_{ij}} \right)^{12} - \left(\frac{\sigma}{d_{ij}} \right)^6 + \frac{1}{4} \right), & 1 \leq i + 1 < j \leq N \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

where d_{ij} is the Euclidean distance between r_i and r_j , $I(*)$ indicates the indicative function, and σ is the thickness of the segment [16].

With the potential field well defined, the state of polymers can be simulated by molecular dynamics. In real simulations, additional constraints may be applied. For example, α and β globin loci represent a classical example of chromatin interaction [17]. To model the interaction, Brackley *et al.* have taken the DNase1 hypersensitive sites (DHSs) as common proxy for binding of proteins and modelled the sites of CCCTC-binding factor (CTCF) and histone modification H3K4me3 separately [18]. On a larger scale, the decondensation (i.e., loosening) of a small chromosome has been simulated by a bead spring polymer model with bead diameter of 30 nm [19]. In the simulation, either linear or ring-shaped helical structures were taken as the initial models, and the effects of linear connectivity, self-avoidance and bending stiffness of the chromatin fiber were considered, while effects from transcription or topoisomerase were neglected. Compared with FISH data in yeast, fly and human, the model has successfully reproduced territory shapes and spatial distances between chromatin loci in the genomes [19]. A whole chromosome can also be simulated by polymer dynamics. For instance, a simulation for the chromosome conformation of interphase haploid budding yeast has been done with Brownian dynamics [20].

POLYMER MODELS FOR LARGE CHROMOSOMES

For larger monomer scales, such as several thousand to several mega bases, the molecular forces neutralize each other, making the chromosomes behave much more like a chain under Brownian motion [21]. Models assume that monomers are flexibly connected to each other. The models are then assessed by certain experimental measurable dimensionless variables, such as the relationship between the spatial (R) and genome (N) distance of two loci. It has been shown that $\langle R^2 \rangle = N^{2\nu}$ [22], where the scaling exponent ν indicates the folding properties of the polymer models. In this category, the random walk model and the self-avoiding walk (SAW) model are fundamental. In the random walk model, the monomers are treated as mathematical points such that multiple points can occupy the same spot simultaneously [22]. In the self-avoiding walk model, a positive volume is assumed for each monomer, implying that the monomers are spatially mutually exclusive [22]. However, both models failed to recapture $R \sim N$ relationship when genome distance > 5 Mb [23]. For example, both models predict that chromosome condensation forms spherical globules, while neither mitotic nor meiotic condensed chromosome condensation is visually spherical [24]. Moreover, the existence of chromatin loops, which is

functionally important, cannot be explained by either model [25,26].

MODELS FOR CHROMATIN LOOPS

To explain chromatin loops, several models have been described in the literature, such as the dynamic loop model [27] (Figure 1A), the random loop model [23] (Figure 1B), the fractal globular model [4], the strings and binders switch (SBS) model [28] (Figure 1C), and the loop extrusion model [29] (Figure 1D). The dynamic loop model was built on top of the random walk model. Given a certain preassigned probability, two randomly walking monomers will be considered to form a stable loop when they meet each other, and a random half-life will also be assigned to the chromatin loop as long as it is formed [27]. Similar to the dynamic loop model, the random loop model is also built on top of the random walk model. In addition to possible loop formation between adjacent monomers, the random loop model also assigns non-adjacent monomer pairs a probability to form loops; as such, loops on all length scales are generated randomly. Both models successfully explained FISH data for the self-organization patterns of chromatin and cell-to-cell variations of the chromatin structure in human [23]. The fractal globular model is based on the so-called “globule”

models, which were originally used to describe a polymer in a poor solvent at equilibrium [4]. The model assumes each monomer to be a globule and iteratively crumples the globules to form globules of globules until only a single globule remains. The resulting structure resembles a Peano curve [4], a continuous fractal trajectory that densely fills 3D space without crossing itself. The model reproduces the relationship between the interaction frequency P_C and the genome distance s of two monomers as $P_c(s) = 1/s^{1.08}$ [30]. In the SBS model, chromosomes are driven by a SAW model in a solution which contains DNA binding factors with a certain concentration [31]. The binding factors are under Brownian motion, and they bind to chromosomes when they are spatially adjacent. A chromatin loop is formed when a factor binds two loci simultaneously. The concentration of the binding factors defines the behavior of the SBS model. With increasing concentration, chromatin conformation varies from the open SAW conformation to a fractal globule-like structure and, finally, a compact, nonfractal conformation, in which ν goes to 0 as genome distance increases, coinciding with experimental observations [32]. Finally, the loop extrusion model considers a specific type of DNA binding protein called motors. A motor is composed of two functional units (heads) which can independently extrude

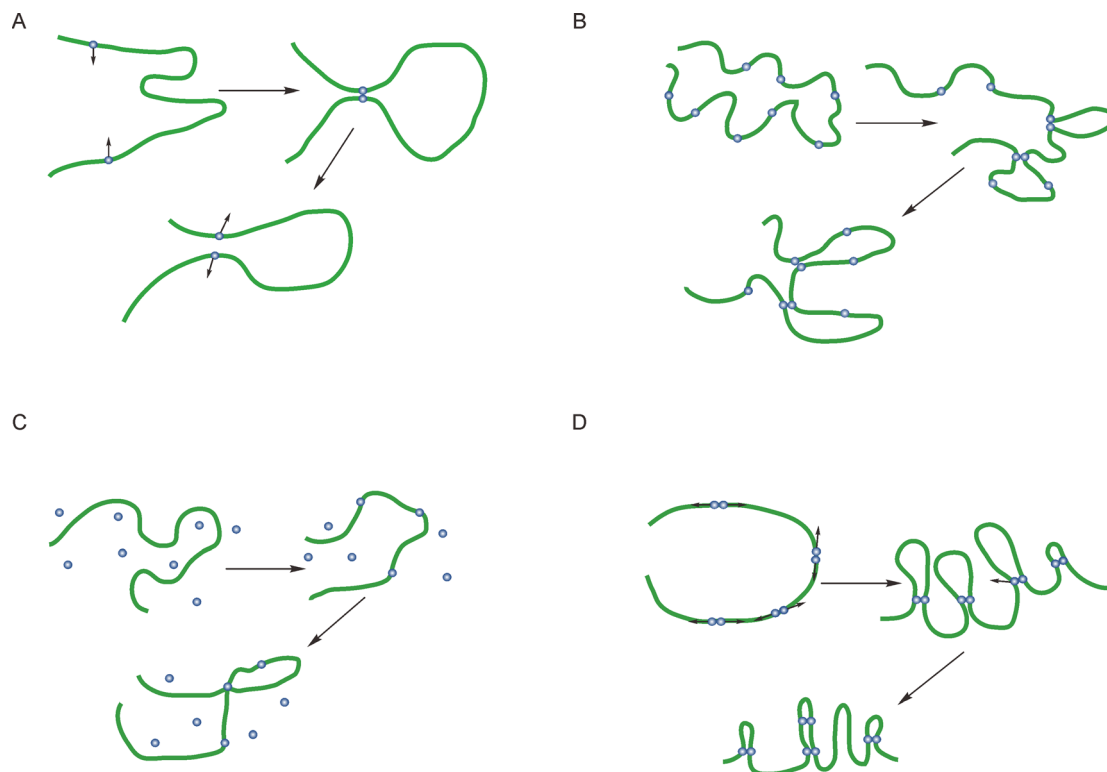


Figure 1. Illustration of polymer models showing the dynamics of chromatin loops. (A) Dynamic loop model. (B) Random loop model. (C) SBS model. (D) Loop extrusion model.

chromatin in opposite directions so that they progressively bridge distal sites. In the loop extrusion model, N motors are assumed to randomly disperse along the chromosome. As long as the two heads of a motor bind to the chromosome, they start to slide away from each other stochastically and extrude a loop. When the heads of two neighboring motors collide, they block each other and halt. Meanwhile, the other heads of each motor remain unperturbed and continue loop extrusion. The motor dissociates from chromatin at a rate r_{off} [29,33]. It has been proposed that protein cohesins and CTCF may act together as the motors in mammals [34]. In addition to reproducing chromatin loops, the loop extrusion model has also been shown to qualitatively and quantitatively reproduce the topologically associated domains (TAD) [35], which explains the mechanism of processive loop enlargement by condensation in mitotic chromosome compaction [33] and chromosome segregation in bacteria [36].

STRUCTURAL MODELS OF THE WHOLE GENOMES BASED ON CHROMOSOME CONFORMATION CAPTURE DATA

With help from next-generation sequencing technology, chromosome conformation capture (3C) [3], together with its variations like 4C [37], 5C [38], Hi-C [4], and ChIA-PET [39], has become a major engine for spatial chromatin structure research [40]. The design principle behind all 3C-based technologies is spatial approximate ligation. More specifically, by ligating DNA fragments that spatially aggregate in nuclear 3D space, information about physical interaction can be converted into DNA sequence arrangement, which can then be examined by PCR(3C) [3], microarray(4C) [37], or next-generation sequencing (5C, Hi-C or ChIA-PET) [4,38,39]. The genome-wide map of chromatin interaction of yeast [41], *Drosophila* [42], mouse [43] and human [4] are all in the public domain. However, with this pairwise chromatin interaction map, one major challenge is to reconstruct the underlying 3D physical model of the genome. Since the initial cell number of nearly all 3C-based technologies is over a million, the interaction map represents all physical structures in the cell population. Given the dynamic nature of chromatin [44], it is hard to distinguish between a stable interaction in a subpopulation and a dynamic interaction in all cells when the observed interaction frequency is not 100% in the map. Thus, two schemes have been devised to model the physical genome structure given the 3C-based contact map. The first scheme aimed to reveal an average, or consensus, structure, and the second scheme models the structures probabilistically, with the ultimate goal of revealing the dynamic of

chromatin in the cell population.

For the first scheme, the common strategy is to seek an optimal structure that best fits the observed chromatin interaction map. The definition of optimal structure completely depends on the objective function which evaluates the goodness of a chromatin structure model representing the experimentally observed interaction frequency (IF) data. Normally, an objective function takes the form $Obj = f(*) + g(*)$, where $f(*)$ denotes the distance between a chromatin structure and IF, and $g(*)$ denotes additional physical constraints. The central part of objective functions is $f(*)$. Since a distance matrix can be conveniently converted into three-dimensional coordinates by multidimensional scaling or semidefinite programming [45,46], $f(*)$ can be roughly written as $f(\mathbf{D}, IF^{-\alpha})$, where \mathbf{D} denotes the distance matrix, and the exponent parameter α can be fixed, or optimized, as a hyperparameter in the model. Depending on $f(\mathbf{D}, IF^{-\alpha})$ and constraints of $g(*)$, various objective functions and optimization algorithms have been proposed [16,41,45–52]. We have already seen that polymer models are variously classified [4,41,47,50]. Taking the fractal globule model as an example, the exponent parameter α is set to be a fixed 1 [4]. Another group of polymer models focus on the design of $g(*)$ to represent physical or experimental constraints, such as the space and shape of cell or nucleus [41,49], as well as such previously noted properties as stretching resistance, excluded volume, and bending rigidity. Many algorithms or platforms have been successfully applied to solve optimal problems, for example including simulated annealing [50], least mean squares (LMS) algorithm [16], shortest-path method [45] or even the Integrative Modeling Platform (IMP), which was originally designed for protein structure prediction [49], or combinations thereof [51]. We recommend a more detailed review of these methods [53].

The second scheme embeds as much accessible information as possible, including observed chromatin interaction map, chromatin physics, and systematic biases of experiments [54], into a probability model and then solves it by either maximum likelihood or the Bayesian approaches. Unlike methods from the first scheme, which only produce average or consensus structure, the models of the second scheme produce an ensemble of structures with likelihood or posterior probability that reflects the dynamics or heterogeneity of chromatin structure in the cell population. Examples of maximum likelihood-based methods include those of Wang *et al.* and Tjong *et al.* [55,56]. In methods based on a specific distribution, e.g., Boltzmann distribution [55] or empirical distribution [56], which take into account physical or biological constraints, the ensemble of structures and probabilities which maximize the total likelihood of the system were inferred.

Examples of Bayesian approaches include BASH and MCMC5C, in which physical constraints were considered in prior probabilities to form a Bayesian inference problem, and a Markov chain Monte Carlo (MCMC) sampling algorithm was designed to sample structures from the posterior distribution [57,58]. The full list of currently available tools can be found in this review [53].

PREDICTING GENOME STRUCTURE BASED ON EPIGENOME DATA

It is well established that epigenetic marks, e.g., the Dnase I hypersensitive site, are strongly associated with chromatin interaction [59] and that CTCF binding is highly enriched at the boundary of topological domains and largely defines them [60,61]. Therefore, researchers ask if 3D chromatin structure could also be predicted from 1D epigenome data, and attempts to do this have been made at different levels. First, global topological genome features, such as compartment partition [62], and TADs [63] can be reconstructed by integration of epigenetic data. Because (i) the two compartments roughly correspond to active and repressive regions in genome and (ii) the association between transcriptional activity of a genome region and its epigenetic state is well established [64–66], it is natural to predict compartments [67]. For TAD prediction, histone markers, together with DNA sequence information, such as PhastCons conservation score, TSS proximity and GC content, and binding profile of CTCF, are integrated into a Bayesian Additive Regression Trees (BART) model, which averages results from an ensemble of regression trees. The authors showed distinct histone modification combinations between chromatin interaction hubs and chromatin regions that do not interact, as well as between TAD boundaries and domain bodies [63]. The authors also integrated Hi-C data, which in turn helped to predict chromatin interaction hubs [63]. A more challenging task is to predict stable chromatin interactions, i.e., chromatin loops from epigenetic data, but three recent works demonstrated that it is possible [68–70]. Thanks to multi-omics data projects, such as ENCODE [71], hundreds of high-throughput genomic and epigenomic data have become publicly available. Chen *et al.* decomposed the correlation matrix of epigenetic profiles and the Hi-C contact matrix using wavelet analysis, while Zhu *et al.* decomposed the matrix of multi-omics profiles by terser decomposition. By decomposition, large-scale structure of the human genome, such as TAD, as well as chromatin-chromatin interaction, can be predicted with high resolution [68,69]. Whale *et al.* developed TargetFinder to predict enhancer-promoter interactions across multiple cell types at fine scale. Almost no single genomic or epigenetic feature has sufficient efficiency on enhancer-promoter interaction

prediction; therefore, the author concluded that complex, but consistent, combinations of genomic and epigenetic features encode the 3D structure of regulatory interactions [70].

FUTURE DIRECTIONS

Although methods and models have successfully revealed features of 3D genome at various scales, the main challenges in the field remain elusive. First, given the current incompatibility of genome size and limited computational power, coarse-graining is a commonly used strategy. Here, monomers in the models could represent genome regions at scales from single nucleosome to megabase length TADs. Thus, the parameters at each level, i.e., energy functions, used in the simulations should be carefully considered. In many cases, however, such parameters still need to be estimated. Second, chromatin structures are believed to be highly dynamic and heterogeneous in the cell population [72]. Although several attempts have been made to model chromatin structure ensembles at given loci [18,73,74], such as globin [74] and Xist [74], or globally by Bayesian approaches, as discussed above [53,57,58], most current models infer an average structure of millions of cells from which the data were generated. Currently, making a precise map of genome structure ensemble is unattainable. Therefore, new computational methods, as well as experimental technologies, that aim to retrieve genome structure information in single cells must be developed. Third, 3D genome structure is always changing along a time scale that includes gene expression regulation, cell cycle, cell differentiation, aging and speciation. Apart from mechanistic considerations, current models have focused on molecular dynamics that take into account a longer span of time. It will be mutually beneficial to evolutionary research of 3D genome and developers of powerful computational methods to understand the 3D genome structure.

ACKNOWLEDGEMENTS

We thank Mr. David Martin for the language correction on the manuscript. This work was supported by grants from the National Nature Science Foundation of China (NSFC, 91540114 and 31271398), the National High Technology Development 863 Program of China (2014AA021103), and Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase) to ZZ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Bingxiang Xu and Zhihua Zhang declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

- Dillon, N. (2008) The impact of gene location in the nucleus on transcriptional regulation. *Dev. Cell*, 15, 182–186
- Miele, A. and Dekker, J. (2008) Long-range chromosomal interactions and gene regulation. *Mol. Biosyst.*, 4, 1046–1057
- Dekker, J., Rippe, K., Dekker, M., Kleckner, N. (2002) Capturing chromosome conformation. *Science*, 295, 1306–1311
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326, 289–293
- Cremer, M., Grasser, F., Lanctôt, C., Müller, S., Neusser, M., Zinner, R., Solovei, I. and Cremer, T. (2008) Multicolor 3D fluorescence *in situ* hybridization for imaging interphase chromosomes. In *The Nucleus*, Hancock. R. Ed. 463, 205–239, Germany: Springer
- Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R. M., Zhu, P. and Li, G. (2014) Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science*, 344, 376–380
- Zhu, P. and Li, G. (2016) Structural insights of nucleosome and the 30-nm chromatin fiber. *Curr. Opin. Struct. Biol.*, 36, 106–115
- Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A. and Dekker, J. (2013) Organization of the mitotic chromosome. *Science*, 342, 948–953
- Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Rusczycki, B., *et al.* (2015) CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, 163, 1611–1627
- Bickmore, W. A. (2013) The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.*, 14, 67–84
- Selvaraj, S., Dixon, J., Bansal, V. and Ren, B. (2013) Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.*, 31, 1111–1118
- Klenin, K., Merlitz, H. and Langowski, J. (1998) A Brownian dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.*, 74, 780–788
- Bednar, J., Furrer, P., Stasiak, A., Dubochet, J., Egelman, E. H. and Bates, A. D. (1994) The twist, writhe and overall shape of supercoiled DNA change during counterion-induced transition from a loosely to a tightly interwound superhelix: possible implications for DNA structure *in vivo*. *J. Mol. Biol.*, 235, 825–847
- Grnbech-Jensen, N., Mashl, R. J., Bruinsma, R. F. and Gelbart, W. M. (1997) Counterion-induced attraction between rigid polyelectrolytes. *Phys. Rev. Lett.*, 78, 2477–2480
- Langowski, J. and Heermann, D. W. (2007) Computational modeling of the chromatin fiber. *Semin. Cell. Dev. Biol.*, 235, 659–667
- Meluzzi, D. and Arya, G. (2013) Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Res.*, 41, 63–75
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002) Looping and interaction between hypersensitive sites in the active β -globin locus. *Mol. Cell*, 10, 1453–1465
- Brackley, C. A., Brown, J. M., Waithe, D., Babbs, C., Davies, J., Hughes, J. R., Buckle, V. J. and Marenduzzo, D. (2016) Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol.*, 17, 59
- Rosa, A. and Everaers, R. (2008) Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.*, 4, e1000153
- Tokuda, N., Terada, T. P. and Sasai, M. (2012) Dynamical modeling of three-dimensional genome organization in interphase budding yeast. *Biophys. J.*, 102, 296–304
- Fudenberg, G. and Mirny, L. A. (2012) Higher-order chromatin structure: bridging physics and biology. *Curr. Opin. Genet. Dev.*, 22, 115–124
- Tark-Dame, M., van Driel, R. and Heermann, D. W. (2011) Chromatin folding—from biology to polymer models and back. *J. Cell Sci.*, 124, 839–845
- Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E. M., Verschure, P. J., Indemans, M. H., Gierman, H. J., Heermann, D. W., van Driel, R., *et al.* (2009) Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl. Acad. Sci. USA*, 106, 3812–3817
- Marko, J. F. and Siggia, E. D. (1997) Polymer models of meiotic and mitotic chromosomes. *Mol. Biol. Cell*, 8, 2217–2231
- Göndör, A. and Ohlsson, R. (2009) Chromosome crosstalk in three dimensions. *Nature*, 461, 212–217
- Kadauke, S. and Blobel, G. A. (2009) Chromatin loops in gene regulation. *Biochim. Biophys. Acta*, 1789, 17–25
- Bohn, M. and Heermann, D. W. (2010) Diffusion-driven looping provides a consistent framework for chromatin organization. *PLoS One*, 5, e12218
- Nicodemi, M., Panning, B. and Prisco, A. (2008) A thermodynamic switch for chromosome colocalization. *Genetics*, 179, 717–721
- Alipour, E. and Marko, J. F. (2012) Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.*, 40, 11202–11212
- Grosberg, A. I., Nechaev, S. K. and Shakhnovich, E. I. (1988) The role of topological limitations in the kinetics of homopolymer collapse and self-assembly of biopolymers. *Biofizika*, 33, 247–253
- Nicodemi, M., Panning, B. and Prisco, A. (2008) A thermodynamic switch for chromosome colocalization. *Genetics*, 179, 717–721
- 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. *Proc. Natl. Acad. Sci. USA*, 109, 16173–16178
- Goloborodko, A., Marko, J. F. and Mirny, L. A. (2016) Chromosome compaction by active loop extrusion. *Biophys. J.*, 110, 2162–2168
- Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159, 1665–1680
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., Mimny, L. A. (2015) Formation of chromosomal domains by loop extrusion. *Cell Rep.*, 15, 2038–2049
- Gruber, S. (2014) Multilayer chromosome organization through DNA bending, bridging and extrusion. *Curr. Opin. Microbiol.*, 22, 102–110
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006) Nuclear organization of

- active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.*, 38, 1348–1354
38. Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., *et al.* (2006) Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.*, 16, 1299–1309
 39. Zhang, J., Poh, H. M., Peh, S. Q., Sia, Y. Y., Li, G., Mulawadi, F. H., Goh, Y., Fullwood, M. J., Sung, W. K., Ruan, X., *et al.* (2012) ChIA-PET analysis of transcriptional chromatin interactions. *Methods*, 58, 289–299
 40. Denker, A. and de Laat, W. (2016) The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.*, 30, 1357–1382
 41. 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, 363–367
 42. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, 148, 458–472
 43. Zhang, Y., McCord, R. P., Ho, Y. J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., Becker, M. S., Alt, F. W. and Dekker, J. (2012) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, 148, 908–921
 44. Hübner, M. R. and Spector, D. L. (2010) Chromatin dynamics. *Annu. Rev. Biophys.*, 39, 471–489
 45. Lesne, A., Riposo, J., Roger, P., Courmac, A. and Mozziconacci, J. (2014) 3D genome reconstruction from chromosomal contacts. *Nat. Methods*, 11, 1141–1143
 46. Zhang, Z., Li, G., Toh, K. C. and Sung, W. K. (2013) 3D chromosome modeling with semi-definite programming and Hi-C data. *J. Comput. Biol.*, 20, 831–846
 47. Varoquaux, N., Ay, F., Noble, W. S. and Vert, J. P. (2014) A statistical approach for inferring the 3D structure of the genome. *Bioinformatics*, 30, i26–i33
 48. Peng, C., Fu, L. Y., Dong, P. F., Deng, Z. L., Li, J. X., Wang, X. T. and Zhang, H. Y. (2013) The sequencing bias relaxed characteristics of Hi-C derived data and implications for chromatin 3D modeling. *Nucleic Acids Res.*, 41, e183
 49. Ba, D. and Marti-Renom, M. A. (2012) Genome structure determination via 3C-based data integration by the Integrative Modeling Platform. *Methods*, 58, 300–306
 50. Zou, C., Zhang, Y. and Ouyang, Z. (2016) HSA: integrating multi-track Hi-C data for genome-scale reconstruction of 3D chromatin structure. *Genome Biol.*, 17, 40
 51. Nowotny, J., Ahmed, S., Xu, L., Oluwadare, O., Chen, H., Hensley, N., Trieu, T., Cao, R. and Cheng, J. (2015) Iterative reconstruction of three-dimensional models of human chromosomes from chromosomal contact data. *BMC Bioinformatics*, 16, 338
 52. Trieu, T. and Cheng, J. (2014) Large-scale reconstruction of 3D structures of human chromosomes from chromosomal contact data. *Nucleic Acids Res.*, 42, e52
 53. Serra, F., Di Stefano, M., Spill, Y. G., Cuartero, Y., Goodstadt, M., Ba, D. and Marti-Renom, M. A. (2015) Restraint-based three-dimensional modeling of genomes and genomic domains. *FEBS Lett.*, 589, 2987–2995
 54. Yaffe, E. and Tanay, A. (2011) Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.*, 43, 1059–1065
 55. Wang, S., Xu, J. and Zeng, J. (2015) Inferential modeling of 3D chromatin structure. *Nucleic Acids Res.*, 43, e54
 56. Tjong, H., Li, W., Kalhor, R., Dai, C., Hao, S., Gong, K., Zhou, Y., Li, H., Zhou, X. J., Le Gros, M. A., *et al.* (2016) Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl. Acad. Sci. USA*, 113, E1663–E1672
 57. Rousseau, M., Fraser, J., Ferraiuolo, M. A., Dostie, J. and Blanchette, M. (2011) Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinformatics*, 12, 414
 58. Hu, M., Deng, K., Qin, Z., Dixon, J., Selvaraj, S., Fang, J., Ren, B. and Liu, J. S. (2013) Bayesian inference of spatial organizations of chromosomes. *PLoS Comput. Biol.*, 9, e1002893
 59. He, C., Wang, X. and Zhang, M. Q. (2014) Nucleosome eviction and multiple co-factor binding predict estrogen-receptor-alpha-associated long-range interactions. *Nucleic Acids Res.*, 42, 6935–6944
 60. Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485, 376–380
 61. Ho, J. W., Jung, Y. L., Liu, T., Alver, B. H., Lee, S., Ikegami, K., Sohn, K. A., Minoda, A., Tolstorukov, M. Y., Appert, A., *et al.* (2014) Comparative analysis of metazoan chromatin organization. *Nature*, 512, 449–452
 62. Fortin, J. P. and Hansen, K. D. (2015) Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol.*, 16, 180
 63. Huang, J., Marco, E., Pinello, L. and Yuan, G. C. (2015) Predicting chromatin organization using histone marks. *Genome Biol.*, 16, 162
 64. Zhang, Z. and Zhang, M. Q. (2011) Histone modification profiles are predictive for tissue/cell-type specific expression of both protein-coding and microRNA genes. *BMC Bioinformatics*, 12, 155
 65. Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Peng, W., Zhang, M. Q., *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, 40, 897–903
 66. Karlič, R., Chung, H. R., Lasserre, J., Vlahovicek, K. and Vingron, M. (2010) Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci. USA*, 107, 2926–2931
 67. Allis, C. D., Jenuwein, T., Reinberg, D., Caparros, M. (2015) *Epigenetics*. New York: Cold Spring Harbor Laboratory Press
 68. Zhu, Y., Chen, Z., Zhang, K., Wang, M., Medovoy, D., Whitaker, J. W., Ding, B., Li, N., Zheng, L. and Wang, W. (2016) Constructing 3D interaction maps from 1D epigenomes. *Nat. Commun.*, 7, 10812
 69. Chen, Y., Wang, Y., Xuan, Z., Chen, M. and Zhang, M. Q. (2016) *De novo* deciphering three-dimensional chromatin interaction and topological domains by wavelet transformation of epigenetic profiles. *Nucleic Acids Res.*, 44, e106
 70. Whalen, S., Truty, R. M. and Pollard, K. S. (2016) Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.*, 48, 488–496
 71. The ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57–74
 72. Kornberg, R. D. and Stryer, L. (1988) Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic*

- Acids Res., 16, 6677–6690
73. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. and Chen, L. (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.*, 30, 90–98
74. Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., Tiana, G. and Heard, E. (2014) Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, 157, 950–963