Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Uncovering the statistical physics of 3D chromosomal organization using data-driven modeling

Vinícius G. Contessoto[1,a], Ryan R. Cheng[1,a] and José N. Onuchic[1,2,3,4]

Abstract

In recent years, much effort has been devoted to understanding the three-dimensional (3D) organization of the genome and how genomic structure mediates nuclear function. The development of experimental techniques that combine DNA proximity ligation with high-throughput sequencing, such as Hi-C, have substantially improved our knowledge about chromatin organization. Numerous experimental advancements, not only utilizing DNA proximity ligation but also high-resolution genome imaging (DNA tracing), have required theoretical modeling to determine the structural ensembles consistent with such data. These 3D polymer models of the genome provide an understanding of the physical mechanisms governing genome architecture. Here, we present an overview of the recent advances in modeling the ensemble of 3D chromosomal structures by employing the maximum entropy approach combined with polymer physics. Particularly, we discuss the minimal chromatin model (MiChroM) along with the "maximum entropy genomic annotations from biomarkers associated with structural ensembles" (MEGABASE) model, which have been remarkably successful in the accurate modeling of chromosomes consistent with both Hi-C and DNA-tracing data.

Addresses

[1] Center for Theoretical Biological Physics, Rice University, Houston, TX, USA
[2] Department of Physics and Astronomy, Rice University, Houston, TX, USA
[3] Department of Chemistry, Rice University, Houston, TX, USA
[4] Department of Biosciences, Rice University, Houston, TX, USA

Corresponding author: Onuchic, José N (jonuchic@rice.edu)
(Contessoto V.G.),
(Cheng R.R.)
[a] Authors contributed equally.

Current Opinion in Structural Biology 2022, **75**:102418

This review comes from a themed issue on **Biophysical Methods**

Edited by **David Fushman** and **Dagmar Ringe**

For complete overview of the section, please refer the article collection - Biophysical Methods

Available online 12 July 2022

https://doi.org/10.1016/j.sbi.2022.102418

Introduction

Since the human DNA sequence was decoded [1] in the early 2000s, significant efforts have been made to understand the three-dimensional (3D) organization of the genome and how its structure is involved in the mediation of nuclear function. Approximately 2 m of DNA decorated with proteins and RNA, collectively called chromatin, is contained within the nucleus of human cells, which is just a few microns in diameter. *In vivo*, the genome structural organization varies in different cell types and these differences play a central role in gene regulation and cell differentiation [2,3]. In addition, interphase chromosomes in each cell type also show large structural variability and are better represented by an structural ensemble instead of a single dominant native structure. Despite this structural diversity, perturbations to these structural ensembles nevertheless affect cellular function and may lead to disease [4–7].

In the past decades, DNA-DNA ligation experiments have opened the way to a systematic study of the genome architecture [8–22] by reporting the frequency that any pair of chromatin loci are observed to be in spatial proximity. High throughput ligation assays, such as Hi-C [12,23], revealed the existence of chromosome territories in human cells and an overall genome organization described by two major compartments called A and B, which can be further divided into subcompartments [23]. Compartment A is correlated with euchromatin and the presence of genes whereas compartment B is generally associated with heterochromatin and transcriptional inactivity. Another key organizational feature observed in these experiments are CTCF-mediated loops, which are generated through a motor-driven process called loop extrusion [23–25]. These loops frequently link promoters and enhancers, indicating their key role in gene expression [26,27].

While Hi-C and related approaches have clearly revealed important aspects of genome organization, determination of the 3D structural ensembles consistent with Hi-C required the development of appropriate modeling. Such models are not only necessary for providing a 3D representation of the genome but also to discover

the underlying mechanisms responsible for genome organization. Successful prediction of these 3D structural ensembles could shed light not only on the genome structural features and its degree of variability for different cell types but could also provide the connection between genome structure and function (e.g., transcription [28]).

Bridging experimental data with theoretical polymer models is, however, challenging. In recent years, numerous polymer models have been used to describe genome organization in three dimensions [29–37, 25,38–44]. In addition to providing 3D structures of chromatin segments, these models also give insights into the physical properties and mechanisms that guide chromatin organization. Simulations employing data-driven models or physics-based approaches reproduce features observed in Hi-C maps (or other chromosome conformation capture techniques), such as topologically associating domains (TADs) [34–37], loop domains [25,38,39], and compartments [40–44]. For example, simulations help to better understand how the epigenome affects the phase separation of chromatin within the inverted and conventional nuclei [45] and how the motor activity of SMC complexes performs loop extrusion of chromatin [25,24]. There are a variety of theoretical models for describing chromatin dynamics and function.

Here we focus on a particular set of data-driven models for modeling chromosomes based on the maximum entropy principle (MaxEnt) [46–48]. This approach is the most straightforward way to create a least biased model that is consistent with experimental observations. In other words, when creating an energy function, once the optimization of the parameters/weights is converged in accordance with experimental constraints, the generated models should faithfully reproduce the experimental observables [49–53,47,48,54,55]. MaxEnt models combined with polymer physics have been shown to accurately capture and predict the genome organization within human cells [56–62,55]. Some approaches employing MaxEnt perform a direct inversion of the Hi-C maps [63,56]. These approaches model the interaction between chromatin loci i and j in a manner that is consistent with the contact frequencies observed in Hi-C experiment while making no assumptions regarding the nature of compartmentalization (e.g., A/B types).

We focus on one particular MaxEnt-derived polymer model called the minimal chromatin model (MiChroM) [57]. MiChroM describes compartmentalization observed in Hi-C maps as the phase separation of biochemically distinct chromatin types. We will also briefly discuss the MaxEnt-based approach called MEGABASE (maximum entropy genomic annotations from biomarkers associated with structural ensembles) [62], which can be used to learn the correlative relationships between the presence of histone modifications along a particular stretch of chromatin and the likelihood of that such stretch being in a particular sub-compartment. Combining both approaches (MEGABASE and MiChroM) led to the creation of a computational pipeline that takes epigenetic information (e.g., histone modification tracks) as input and generates the 3D chromosomal structural ensembles of individual chromosomal territories as output [62,58].

In this review, we will discuss recent developments in the physical modeling of chromosomes using MEGABASE and MiChroM and the validation of the simulated structures by comparing them to experimental structures obtained using microscopy (DNA tracing) [61,64–69]. We will further discuss our perspectives on the strengths and limitations of these structural models and make suggestion and predictions about the future of computational genome modeling.

Modeling chromosomes by integrating theory and experiments

The minimal chromatin model (MiChroM)

The minimal chromatin model (MiChroM) [57] is a coarse-grained polymer model of individual chromosomes that was trained on Hi-C data of human lymphoblastoid cells (GM12878) [23] using the principle of maximum entropy. This approach has been widely used for modeling genome organization [57,62,70,58–61]. The MiChroM energy function was conceived based on three physical assumptions. The first assumption is that the compartmentalization observed in the DNA-DNA ligation maps result from the phase separation of chromatin loci of distinct type. These chromatin types were called A1, A2, B1, B2, and B3, which give rise to the corresponding sub-compartments observed in experiment [23]; an additional non-specific chromatin type called NA was also modeled. The second assumption is that certain pairs of loci (separated by kilobases to megabases) form the anchor of a chromatin loop. It is generally understood that these loops are formed by the structural maintenance of chromosomes (SMC) protein complexes, which extrude a loop until reaching a convergent pair of CCCTC-binding factor (CTCF)–binding motifs [23,24,71]. The final assumption is that for every pair of chromatin loci that comes into spatial proximity, there is arelative gain or loss of free energy $\gamma(d)$ that depends on the genomic distance d that separates the two loci. This assumption is referred to as the ideal chromosome (IC) [57], capturing the translationally invariant local structure of chromatin that exists even in the absence of compartmentalization (phase separation) and looping interactions.

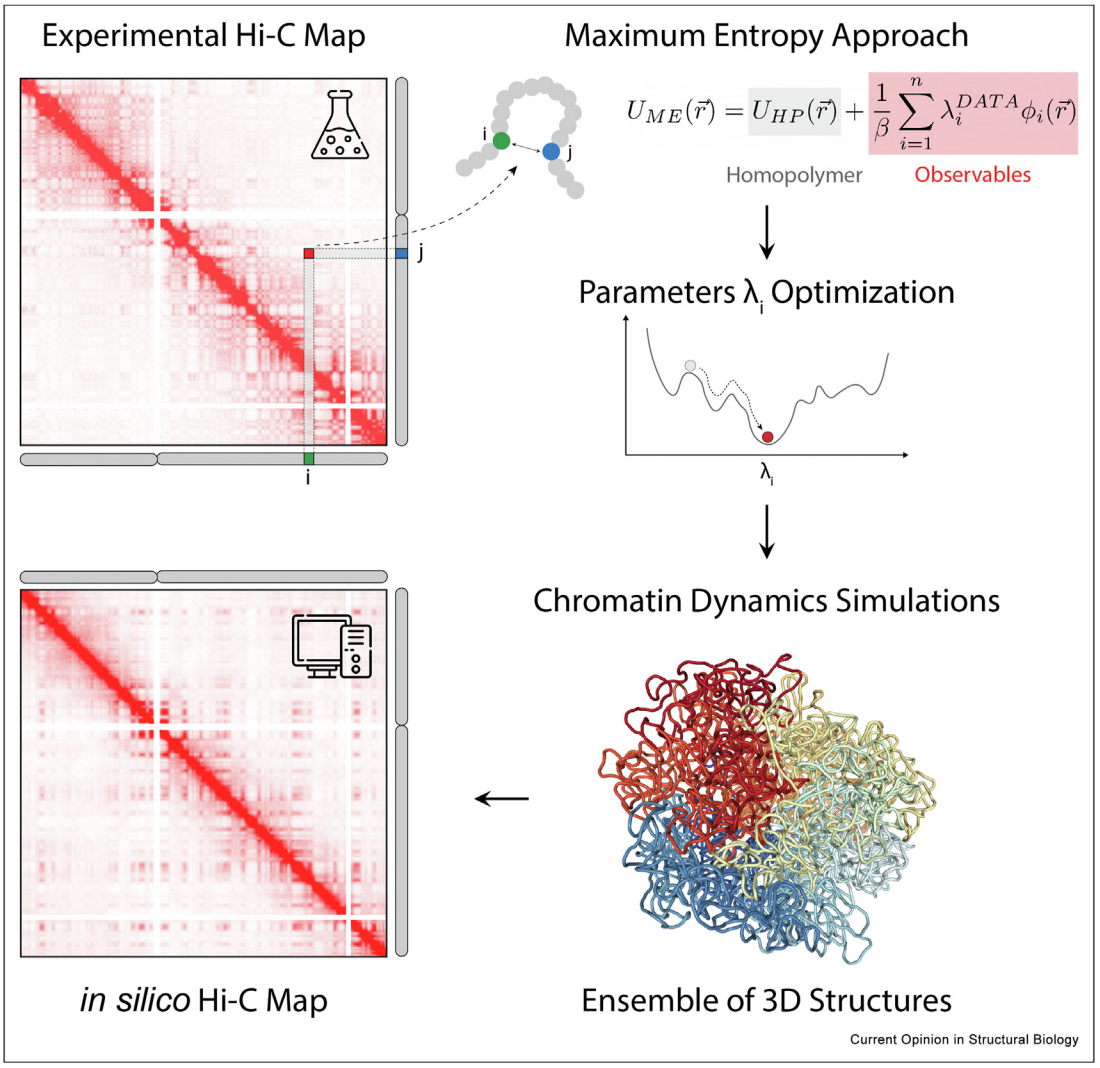
The energy function of MiChroM is given by:

$$U_{\mathrm{MiChroM}}(\vec{r}) = U_{\mathrm{HP}}(\vec{r}) + \sum_{\substack{k \geq l \\ k,l \in \mathrm{Types}}} \alpha_{kl} \sum_{\substack{i \in \{\mathrm{Loci\ of\ Type\ k}\} \\ j \in \{\mathrm{Loci\ of\ Type\ l}\}}} f(r_{ij}) + \chi \sum_{(i,j) \in \{\mathrm{Loop\ Sites}\}} f(r_{ij}) + \sum_{d=3}^{d_{\mathrm{cutoff}}} \gamma(d) \sum_{i} f(r_{i,i+d}) \quad (1)$$

where $f(r_{ij}) = \frac{1}{2}\left(1 + \tanh\left[\mu\left(r_c - r_{ij}\right]\right)\right)$ describes the probability of crosslinking between chromatin loci i and j separated by a distance r_{ij}. The $U_{\mathrm{HP}}(\vec{r})$ describes the connectivity of the polymer model while the remaining terms with parameters α_{kl}, χ, and $\gamma(d)$ describe the phase separation (compartmentalization), looping interactions, and ideal chromosome, respectively. Obtaining the parameters α_{kl}, χ, and $\gamma(d)$ requires an iterative optimization procedure to produce a simulated DNA-DNA ligation map that is consistent with the experimental map.

Figure 1 presents a schematic description of the parameter optimization procedure for modeling chromosomes. As a validation of the training procedure, some features of the *in silico* Hi-C map obtained from the

Figure 1



Workflow for employing the maximum entropy approach and parameter optimization for modeling 3D chromosomal structures. An experimental Hi-C map is used as input data. The observable parameter is associated with the loci contact frequency, that is, the higher the loci pair contact frequency, stronger is the interaction between a pair of chromatin loci in the polymer model. An iterative training procedure determines these parameter values (See [57] for additional details). Once the optimal parameters are obtained, they are used in the energy function for the production simulations. These simulations generate an ensemble of 3D structures. The average of the contact probability over this ensemble of chromosomal structures presents an *in silico* Hi-C map consistent with the experiments.

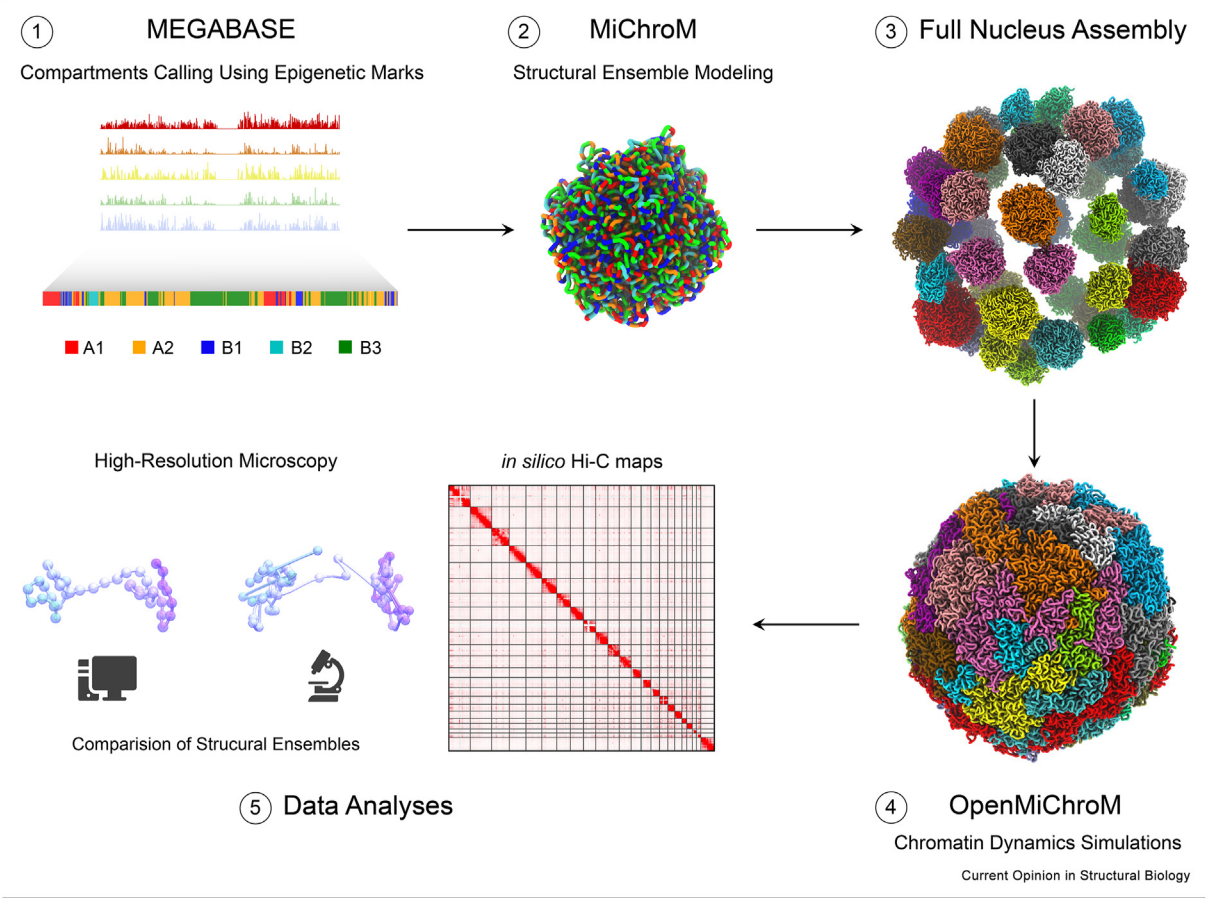
optimized energy function are compared to the experimental Hi-C map. These features include the contact probability decay curve (polymer length scaling), the eigenvector extracted from the correlation matrix (A/B calling), correlation of the probabilities as a function of the genomic separation, etc. Refer to the study by Di Pierro et al. [57] for additional details of the MiChroM potential as well as details of the parameter optimization and validation.

MaxEnt models employed for chromosome modeling are considered a top-down approach in which the physical interactions are trained and calibrated to reproduce experimental observables, such as the Hi-C data [57,59,58,56]. The value of these interactions is fitted to experimental data and therefore not computed from first principles. On the other hand, untrained bottom-up models can nevertheless be used to explore underlying mechanisms of genome organization [40,72].

MEGABASE: Learning the correlative relationships between epigenetic modifications and compartmentalization

The MiChroM prediction of a 3D chromosomal structural ensemble given a sequence of chromatin subtypes (A1, A2, B1, B2, and B3) is in some sense analogous to the prediction of a protein fold given a sequence of amino acids [73]. It was subsequently shown that a neural

Figure 2



MEGABASE + MiChroM computational pipeline. MEGABASE uses as input epigenetic modification information (such as ChIP-seq tracks) to predict the chromatin structural type (associated with sub-compartmentalization) for each locus. This sequence of chromatin types is loaded into the MiChroM energy function which generates an ensemble of 3D structures. This predicted structural ensemble has been shown to be consistent with experimental data from Hi-C and FISH [23,62]. The chromatin dynamics simulations and analyses are performed using the OpenMiChroM software package. The OpenMiChroM platform performs chromosome simulations using GPUs via OpenMM Python API which allows the investigation of large systems, such as the full nucleus of a human cell at 50 kilobase resolution. This approach was also successful in predicting the genome organization of multiple human cell lines, capturing the high degree of structural variability of chromosomes in the interphase [61] consistent with DNA tracing [64]. An example of an open dumbbell-like structure from simulation and experimental tracing, respectively, is shown in the bottom left. The MEGABASE + MiChroM computational pipeline is available at the Nucleome Data Bank (NDB) [58], a web platform that stores 3D chromosomal structures from experiments (high-resolution microscopy) and modeling and it is freely and publicly available at https://ndb.rice.edu.

network-based approach called MEGABASE [62] can quantify the correlative relationship between the presence of particular histone modifications along a segment of chromatin and the sub-compartmentalization of that segment [23]. The inferred model of MEGABASE is a statistical model of state vectors characterizing each locus of chromatin l: $\sigma(l) = (C(l), Exp_1(l), \ldots, Exp_L(l))$, where C is the sub-compartment annotation of locus l from Hi-C experiments [23] and Exp_i denotes the discretized chromatin immunoprecipitation (ChIP-Seq) signal from experiment for a particular histone modification track. The resultant model allows for the inference of the chromatin structural types $CST(l)$ (compartment annotation) given experimental data for histone modifications:

$$CST(l) = argmax\ P\left(C \middle| Exp_{1,\ldots,L}(l-2, l-1, l, l+1, l+2)\right) \quad (2)$$

In practice, the experimental ChIP-Seq signals from adjacent chromatin loci $l-2$, $l-1$, $l+1$, and $l+2$ are used to predict the chromatin type of locus l. Figure 2 gives a schematic representation of the MEGABASE + MiChroM computational framework. Refer to Ref: [62] for additional details of the MEGABASE as well as details of the parameter optimization.

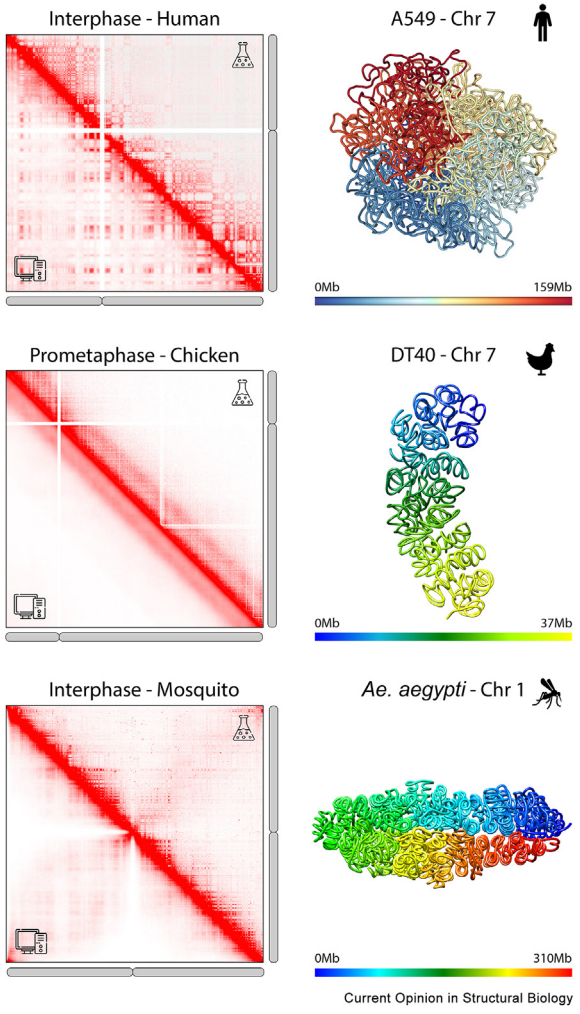
OpenMiChroM: A fast and scalable platform for genome simulations

Another recent tool released for chromosome modeling is the software package OpenMiChroM, a fast and scalable platform that performs chromatin simulations using GPUs via OpenMM Python API [74]. OpenMiChroM chromatin dynamics simulations allow for the investigation of larger systems, such as the full nucleus of a human cell that has around 200,000 chromatin beads at 50 kilobase resolution. In addition, OpenMiChroM generates chromosome trajectories stored in a *cndb* file format (binary version of the *ndb* file) and also provides several built-in analysis tools. OpenMiChroM installation guide and tutorials can be found at https://openmichrom.readthedocs.io. OpenMiChroM also performs training and parameters optimization using first- or second-order minimization methods.

As previously mentioned, the training of the MiChroM energy function involves the optimization of a set of parameters α_{kl}, χ, and $\gamma(d)$ that describe the phase separation, looping interactions, and the ideal chromosome term. The phase separation parameters, α_{kl}, represent the energetic stabilization between loci of chromatin type k and l when they are spatially proximal. There are 21 different values of α_{kl} corresponding to the interaction between the 6 chromatin types described in MiChroM (A1, A2, B1, B2, B3, and NA). A single value of χ is trained to describe the energetic stabilization between chromatin loci that form a loop anchor. The $\gamma(d)$ parameter describes the ideal chromosome that varies depending on the genomic separation between a pair of chromatin loci, d. A value of $\gamma(d)$ is trained for each fixed value of d up to separations of hundreds of beads. Using OpenMiChroM to reproduce MiChroM training, the process takes a few hours if the simulations are performed using a single desktop GPU.

Figure 3 shows examples of optimized energy functions for chromosome modeling of different organisms and cell phases during the cell cycle generated using the

Figure 3



The OpenMiChroM platform can be used for modeling chromosomes of different organisms. OpenMiChroM chromatin dynamics simulations create an ensemble of 3D chromosomal structures [59]. The average of the contact probability over this ensemble generates an *in silico* Hi-C map (lower triangle) consistent with the 2D experimental Hi-C matrix (upper triangle). The modeling can also be performed for different phases of the cell cycle. On The top and bottom shows the human [23] and the mosquito [75,76] maps, respectively. Both are modeled in the interphase. The middle shows the chicken chromosome structure obtained for the prometaphase [38].

OpenMiChroM platform. Refer to Refs: [59,60] for additional details of the OpenMiChroM as well as details of the optimization procedure.

MEGABASE and MiChroM: going from epigenetic data to an ensemble of 3D chromosomal structures

The combined approaches of MEGABASE and MiChroM have been shown to generate a 3D ensemble of chromosomal structures that is consistent with population averaged Hi-C maps for chromosomes in the interphase [62,61,58]. To date, MEGABASE and MiChroM have made successful predictions of 3D structures for many other human cell types that were not considered in their respective training and optimization procedures [61,58]. This supports the hypothesis that it is the phase separation of biochemically distinct segments of chromatin that gives rise to compartmentalization that is observed in DNA-DNA ligation experiments. Further, the distinct biochemical properties of these marked segments of chromatin appear to be directly related to the enrichment of chemical modifications to the histone tails (e.g., methylation or acetylation) along that segment.

Although MiChroM was trained on DNA-DNA ligation maps and not calibrated using any time information, it has nevertheless been able to correctly capture many of the dynamical aspects of the genome. Active processes and motors that act on the genome are not explicitly included but approximated through a quasi-equilibrium energy landscape. In particular, molecular simulations of chromosomes using MiChroM capture the sub-diffusive motion, spatial coherence, and viscoelasticity [70] of chromatin consistent with experimental measurements [77].

Interestingly, early analysis of our simulated chromosomes revealed that they exhibited a liquid-like, heterogeneous structural distribution [56,57]. This high degree of structural heterogeneity was recently directly observed using DNA tracing experiments, in which chromatin structures were imaged using microscopy [61,64–68,57,58,78,69]. It is becoming increasingly clear that the observed structural variability is a hallmark of chromatin in the interphase. A recent analysis [61] comparing chromatin structures from imaging [64] and computer simulation (MEGABASE and MiChroM) revealed structural transitions of a ~ 2 megabase gene-containing segment of chromatin between closed structures and open dumbbell-like structures (See Figure 2. Not only were the structures from experiment and simulation observed to be consistent, but the degree of structural variability was also in quantitative agreement. In particular, the apparent free energy difference between open and closed structures remarkably was found to be $\sim 4k_BT$ in both the experimentally traced [64] and simulated structures. This shows that our computational pipeline also captures the structural variability observed in real chromatin and further demonstrates that chromosomal structures can only truly be described from the energy landscape (ensemble) perspective. Understanding the structural features of this ensemble and its connection to cell function, such as gene expression and cell differentiation, remains a current challenge in genome research.

Perspectives on future genomic modeling

MEGABASE and MiChroM could serve as important tools for studying genome organization in a variety of cell types and tissues. These models can be used to study the global features of genome organization through differentiation and development and can potentially be used to examine structural perturbations associated with certain diseases, such as cancers. MEGABASE and MiChroM can also potentially be used to examine genome architectures across the tree of life; similar tools have been used to explore the emergence of different genome architectures across different organisms [79] or different phases of the cell cycle [38].

While MiChroM appears to accurately capture chromosomal organization at a resolution of 50 kilobases per monomer, there exist several limitations in its current implementation. In reality, the structures and functional aspects of the genome are shaped by the action of various motor proteins and active processes. The molecular details of these interactions become essential as the resolution of the polymer model approaches base pair resolution. In its current formulation, MiChroM describes these active processes using a quasi-equilibrium approximation and coarse-grains the biomolecular interactions that shape the genome in a statistical manner. For example, the motor activity of condensin and cohesin protein complexes are indirectly captured by the ideal chromosome term of the MiChroM energy function (See eq. 1), which describes the genomic distance-dependent compaction and ordering that is attributed to the motor activity of these protein complexes.

In principle, the MiChroM model can be trained and calibrated for a polymer model at a finer resolution than the current one of 50 kilobases per monomeric unit, provided that a sufficiently high resolution DNA-DNA ligation map is available. However, the statistical treatment of biomolecular interactions along the DNA polymer breaks down as the resolution approaches levels of few base pairs. Fine resolution models of chromatin require a detailed treatment of those biomolecules that shape organization at those scales, such as histone complexes, transcriptional machinery, motor proteins, etc. Furthermore, modeling of gene expression regions will also need to include DNA supercoiling effects, which are not accounted for by the current MiChroM model.

Future work where chromatin models are created at a finer resolution can still be constructed with MaxEnt by incorporating finer resolution constraints during optimization. Clearly, MiChroM will need to be revised as we move toward these more detailed models. Future MiChroM-related models can potentially be trained with other data sources, such as the high-resolution chromosome snapshots from DNA tracing.

Acknowledgments

We want to thank Peter G. Wolynes, Michele Di Pierro, Antonio B. Oliveira Junior, Sumitabha Brahmachari, Matheus Mello, and Esteban Dodero Rojas for many useful conversations during the writing of this work and for all of their comments and suggestions. This research was supported by the Center for Theoretical Biological Physics sponsored by the NSF (Grants PHY-2019745 and CHE-1614101) and by the Welch Foundation (Grant C-1792). JNO is a Cancer Prevention and Research Institute of Texas (CPRIT) Scholar in Cancer Research.

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, *et al.*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**:860–921, https://doi.org/10.1038/35057062.

2. Cremer T, Cremer C: **Chromosome territories, nuclear architecture and gene regulation in mammalian cells**. *Nat Rev Genet* 2001, **2**:292–301, https://doi.org/10.1038/35066075.

3. Bickmore WA: **The spatial organization of the human genome**. *Annu Rev Genom Hum Genet* 2013, **14**:67–84, https://doi.org/10.1146/annurev-genom-091212-153515. arXiv:.

4. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, *et al.*: **An oestrogen-receptor-α-bound human chromatin interactome**. *Nature* 2009, **462**:58–64, https://doi.org/10.1038/nature08497.

5. Göndör A: **Dynamic chromatin loops bridge health and disease in the nuclear landscape**. *Semin Cancer Biol* 2013, **23**:90–98, https://doi.org/10.1016/j.semcancer.2013.01.002.

6. Krijger PHL, de Laat W: **Regulation of disease-associated gene expression in the 3D genome**. *Nat Rev Mol Cell Biol* 2016, **17**:771–782, https://doi.org/10.1038/nrm.2016.138.

7. Montefiori L, Wuerffel R, Roqueiro D, Lajoie B, Guo C, Gerasimova T, *et al.*: **Extremely long-range chromatin loops link topological domains to facilitate a diverse antibody repertoire**. *Cell Rep* 2016, **14**:896–906, https://doi.org/10.1016/j.celrep.2015.12.083.

8. Dekker J, Rippe K, Dekker M, Kleckner N: **Capturing chromosome conformation**. *Science* 2002, https://doi.org/10.1126/science.1067799.

9. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, *et al.*: **Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements**. *Genome Res* 2006, **16**:1299–1309, https://doi.org/10.1101/gr.5571506.

10. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, *et al.*: **Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C)**. *Nat Genet* 2006, **38**:1348–1354, https://doi.org/10.1038/ng1896.

11. Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, *et al.*: **Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions**. *Nat Genet* 2006, **38**:1341–1347, https://doi.org/10.1038/ng1891.

12. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, *et al.*: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome**. *Science* 2009, https://doi.org/10.1126/science.1181369.

13. Hakim O, Misteli T: **SnapShot: chromosome conformation capture**. *Cell* 2012, **148**, https://doi.org/10.1016/j.cell.2012.02.019.

14. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, *et al.*: **HiChIP: efficient and sensitive analysis of protein-directed genome architecture**. *Nat Methods* 2016, **13**:919–922, https://doi.org/10.1038/nmeth.3999.

15. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, *et al.*: **Iterative correction of Hi-C data reveals hallmarks of chromosome organization**. *Nat Methods* 2012, **9**:999–1003, https://doi.org/10.1038/nmeth.2148.

16. Olivares-Chauvet P, Mukamel Z, Lifshitz A, Schwartzman O, Elkayam NO, Lubling Y, *et al.*: **Capturing pairwise and multi-way chromosomal conformations using chromosomal walks**. *Nature* 2016, **540**:296–300, https://doi.org/10.1038/nature20158.

17. Beagrie RA, Scialdone A, Schueler M, Kraemer DCA, Chotalia M, Xie SQ, *et al.*: **Complex multi-enhancer contacts captured by genome architecture mapping**. *Nature* 2017, **543**:519–524, https://doi.org/10.1038/nature21411.

18. Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, *et al.*: **Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus**. *Cell* 2018, **174**:744–757, https://doi.org/10.1016/j.cell.2018.05.024. e24.

19. Akgol Oksuz B, Yang L, Abraham S, Venev SV, Krietenstein N, Parsi KM, *et al.*: **Systematic evaluation of chromosome conformation capture assays**. *Nat Methods* 2021, **18**:1046–1055, https://doi.org/10.1038/s41592-021-01248-7.

20. Jerkovic´ I, Cavalli G: **Understanding 3D genome organization by multidisciplinary methods**. *Nat Rev Mol Cell Biol* 2021, **22**:511–528, https://doi.org/10.1038/s41580-021-00362-w.

21. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, *et al.*: **Topological domains in mammalian genomes identified by analysis of chromatin interactions**. *Nature* 2012, **485**:376–380, https://doi.org/10.1038/nature11082.

22. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, *et al.*: **Single-cell Hi-C reveals cell-to-cell variability in chromosome structure**. *Nature* 2013, **502**:59–64, https://doi.org/10.1038/nature12593.

23. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping**. *Cell* 2014, **159**:1665–1680, https://doi.org/10.1016/j.cell.2014.11.021.

24. Sanborn AL, Rao SSP, Huang SC, Durand NC, Huntley MH, Jewett AI, *et al.*: **Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes**. *Proc Natl Acad Sci USA* 2015, **112**:E6456–E6465, https://doi.org/10.1073/pnas.1518552112.

25. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA: **formation of chromosomal domains by loop extrusion**. *Cell Rep* 2016, **15**:2038–2049, https://doi.org/10.1016/j.celrep.2016.04.085.

26. Holwerda S, De Laat W: **Chromatin loops, gene positioning, and gene expression**. *Front Genet* 2012, **3**.

27. Zhao L, Wang S, Cao Z, Ouyang W, Zhang Q, Xie L, *et al.*: **Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation**. *Nat Commun* 2019, **10**:3640, https://doi.org/10.1038/s41467-019-11535-9.

28. Leidescher S, Ribisel J, Ullrich S, Feodorova Y, Hildebrand E, Galitsyna A, *et al.*: **Spatial organization of transcribed eukaryotic genes**. *Nat Cell Biol* 2022, **24**:327–339, https://doi.org/10.1038/s41556-022-00847-6.

29. Münkel C, Langowski J: **Chromosome structure predicted by a polymer model**. *Phys Rev* 1998, **57**:5888–5896, https://doi.org/10.1103/PhysRevE.57.5888.

30. Rosa A, Everaers R: **Structure and dynamics of interphase chromosomes**. *PLoS Comput Biol* 2008, **4**:e1000153, https://doi.org/10.1371/journal.pcbi.1000153.

31. Mateos-Langerak J, Bohn M, de Leeuw W, Giromus O, Manders EMM, Verschure PJ, *et al.*: **Spatially confined folding of chromatin in the interphase nucleus**. *Proc Natl Acad Sci USA* 2009, **106**:3812–3817, https://doi.org/10.1073/pnas.0809501106.

32. Barbieri M, Chotalia M, Fraser J, Lavitas LM, Dostie J, Pombo A, *et al.*: **Complexity of chromatin folding is captured by the strings and binders switch model**. *Proc Natl Acad Sci USA* 2012, **109**:16173–16178, https://doi.org/10.1073/pnas.1204799109.

33. Brahmachari S, Contessoto VG, Di Pierro M, Onuchic JN: **Shaping the genome via lengthwise compaction, phase separation, and lamina adhesion**. *Nucleic Acids Res* 2022: gkac231, https://doi.org/10.1093/nar/gkac231.

34. Michieletto D, Orlandini E, Marenduzzo D: **Polymer model with Epigenetic Recoloring Reveals a Pathway for the de novo Establishment and 3D Organization of Chromatin Domains**. *Phys Rev X* 2016, **6**:41047, https://doi.org/10.1103/PhysRevX.6.041047.

35. Shukron O, Holcman D: **Statistics of randomly cross-linked polymer models to interpret chromatin conformation capture data**. *Phys Rev* 2017, **96**:12503, https://doi.org/10.1103/PhysRevE.96.012503.

36. Sun Q, Perez-Rathke A, Czajkowsky DM, Shao Z, Liang J: **High-resolution single-cell 3D-models of chromatin ensembles during Drosophila embryogenesis**. *Nat Commun* 2021, **12**:205, https://doi.org/10.1038/s41467-020-20490-9.

37. Shukron O, Holcman D: **Transient chromatin properties revealed by polymer models and stochastic simulations constructed from Chromosomal Capture data**. *PLoS Comput Biol* 2017, **13**:e1005469, https://doi.org/10.1371/journal.pcbi.1005469.

38. Gibcus JH, Samejima K, Goloborodko A, Samejima I, Naumova N, Nuebler J, *et al.*: **A pathway for mitotic chromosome formation**. *Science* 2018, **359**:eaao6135, https://doi.org/10.1126/science.aao6135.

39. Brahmachari S, Marko JF: **Chromosome disentanglement driven via optimal compaction of loop-extruded brush structures**. *Proc Natl Acad Sci USA* 2019, **116**:24956–24965, https://doi.org/10.1073/pnas.1906355116.

40. MacPherson Q, Beltran B, Spakowitz AJ: **Bottom–up modeling of chromatin segregation due to epigenetic modifications**. *Proc Natl Acad Sci USA* 2018, **115**:12739–12744, https://doi.org/10.1073/pnas.1812268115.

41. Shi G, Liu L, Hyeon C, Thirumalai D: **Interphase human chromosome exhibits out of equilibrium glassy dynamics**. *Nat Commun* 2018, **9**:3161, https://doi.org/10.1038/s41467-018-05606-6.

42. Brackley CA, Johnson J, Kelly S, Cook PR, Marenduzzo D: **Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains**. *Nucleic Acids Res* 2016, **44**:3503–3512, https://doi.org/10.1093/nar/gkw135.

43. Jost D, Carrivain P, Cavalli G, Vaillant C: **Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains**. *Nucleic Acids Res* 2014, **42**:9553–9561, https://doi.org/10.1093/nar/gku698.

44. Belaghzal H, Borrman T, Stephens AD, Lafontaine DL, Venev SV, Weng Z, *et al.*: **Liquid chromatin Hi-C characterizes compartment-dependent chromatin interaction dynamics**. *Nat Genet* 2021, **53**:367–378, https://doi.org/10.1038/s41588-021-00784-4.

45. Falk M, Feodorova Y, Naumova N, Imakaev M, Lajoie BR, Leonhardt H, *et al.*: **Heterochromatin drives compartmentalization of inverted and conventional nuclei**. *Nature* 2019, **570**:395–399, https://doi.org/10.1038/s41586-019-1275-3.

46. Jaynes ET: **Information theory and statistical mechanics**. *Phys Rev* 1957, **106**:620–630, https://doi.org/10.1103/PhysRev.106.620.

47. Pitera JW, Chodera JD: **On the use of experimental observations to bias simulated ensembles**. *J Chem Theor Comput* 2012, **8**:3445–3451, https://doi.org/10.1021/ct300112v.

48. Roux B, Weare J: **On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method**. *J Chem Phys* 2013, **138**:84107, https://doi.org/10.1063/1.4792208.

49. Seno F, Trovato A, Banavar JR, Maritan A: **Maximum entropy approach for deducing amino acid interactions in proteins**. *Phys Rev Lett* 2008, **100**:78102, https://doi.org/10.1103/PhysRevLett.100.078102.

50. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, *et al.*: **Direct-coupling analysis of residue coevolution captures native contacts across many protein families**. *Proc Natl Acad Sci USA* 2011, **108**:E1293–E1301, https://doi.org/10.1073/pnas.1111471108.

51. Pressé S, Ghosh K, Lee J, Dill KA: **Principles of maximum entropy and maximum caliber in statistical physics**. *Rev Mod Phys* 2013, **85**:1115–1141, https://doi.org/10.1103/RevModPhys.85.1115.

52. Cesari A, Reißer S, Bussi G: **Using the maximum entropy principle to combine simulations and solution experiments**. *Computation* 2018, **6**:15, https://doi.org/10.3390/computation6010015.

53. De Martino A, De Martino D: **An introduction to the maximum entropy approach and its application to inference problems in biology**. *Heliyon* 2018, **4**:e00596, https://doi.org/10.1016/j.heliyon.2018.e00596.

54. Latham AP, Zhang B: **Maximum entropy optimized force field for intrinsically disordered proteins**. *J Chem Theor Comput* 2020, **16**:773–781, https://doi.org/10.1021/acs.jctc.9b00932.

55. • Lin X, Qi Y, Latham AP, Zhang B: **Multiscale modeling of genome organization with maximum entropy optimization**. *J Chem Phys* 2021, **155**:10901, https://doi.org/10.1063/5.0044150.

This work discusses the application of the maximum entropy approach for investigating the genome organization and dynamics in different length scales using coarse-grained models. The work presents some applications, including models at finer resolution that can be used to investigate the phase separation of disordered proteins and dictate chromatin stability in situ.

56. Zhang B, Wolynes PG: **Topology, structures, and energy landscapes of human chromosomes**. *Proc Natl Acad Sci USA* 2015, **112**:6062–6067, https://doi.org/10.1073/pnas.1506257112.

57. Pierro MD, Zhang B, Aiden EL, Wolynes PG, Onuchic JN: **Transferable model for chromosome architecture**. *Proc Natl Acad Sci USA* 2016, **113**:12168–12173, https://doi.org/10.1073/pnas.1613607113.

58. Contessoto VG, Cheng RR, Hajitaheri A, Dodero-Rojas E, Mello MF, Lieberman-Aiden E, *et al.*: **The Nucleome Data Bank: web-based resources to simulate and analyze the three-dimensional genome**. *Nucleic Acids Res* 2021, **49**:D172–D182, https://doi.org/10.1093/nar/gkaa818.

59. • Oliveira Junior AB, Contessoto VG, Mello MF, Onuchic JN: **A scalable computational approach for simulating complexes of multiple chromosomes**. *J Mol Biol* 2021, **433**:166700, https://doi.org/10.1016/j.jmb.2020.10.034.

This work introduces the OpenMiChroM software, a Python library for performing chromatin dynamics simulatidns. OpenMiChroM uses the OpenMM Python API employing the MiChroM (Minimal Chromatin Model) energy function. OpenMiChroM allows simulations of a single or multiple chromosome chain using High-Performance Computing in different platforms (GPUs and CPUs). OpenMiChroM also performs the parameters optimization for simulating the genome of different organisms and phases of the cell cycle.

60. Oliveira Junior AB, Estrada CP, Aiden EL, Contessoto VG, Onuchic JN: **Chromosome Modeling on Downsampled Hi-C Maps Enhances the Compartmentalization Signal**. *J Phys Chem B* 2021, **125**:8757–8767, https://doi.org/10.1021/acs.jpcb.1c04174.

61. • Cheng RR, Contessoto VG, Lieberman Aiden E, Wolynes PG, Di Pierro M, Onuchic JN: **Exploring chromosomal structural heterogeneity across multiple cell lines**. *Elife* 2020, **9**:e60312, https://doi.org/10.7554/eLife.60312.
This work used computer simulations to predict the cell type specific 3D chromosomal structural ensembles for different euman cell types. The predicted structures were found to be in quantitative agreement with both DNA-DNA ligation experiments (Hi-C) as well as recent microscopy experiments (DNA tracing). The simulated chromatin structures captured the same degree of structural variability that was observed in tracing experiments. In particular, a short gene-containing segment of chromatin was found to transition between open dumbbell-like structures and closed structures in both simulation and experiment with the same apparent free energy difference between open and closed structures. This work also highlighted the radial dependence of compartmentalization within a chromosome territory. In simulated chromosomes, inactive chromatin tended toward the interior of the territory while gene-active chromatin tended toward the surface.

62. Pierro MD, Cheng RR, Aiden EL, Wolynes PG, Onuchic JN: **De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture**. *Proc Natl Acad Sci USA* 2017, **114**:12126–12131, https://doi.org/10.1073/pnas.1714980114.

63. Shi G, Thirumalai D: **From Hi-C contact map to three-dimensional organization of interphase human chromosomes**. *Phys Rev X* 2021, **11**:11051, https://doi.org/10.1103/PhysRevX.11.011051.

64. Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, *et al.*: **Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells**. *Science* 2018, **362**:eaau1783, https://doi.org/10.1126/science.aau1783.

65. Nir G, Farabella I, Estrada CP, Ebeling CG, Beliveau BJ, Sasaki HM, *et al.*: **Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling**. *PLoS Genet* 2018, **14**:e1007872, https://doi.org/10.1371/journal.pgen.1007872.

66. Beliveau BJ, Boettiger AN, Avendaño MS, Jungmann R, McCole RB, Joyce EF, *et al.*: **Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes**. *Nat Commun* 2015, **6**:7147, https://doi.org/10.1038/ncomms8147.

67. Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, *et al.*: **Super-resolution imaging reveals distinct chromatin folding for different epigenetic states**. *Nature* 2016, **529**:418–422, https://doi.org/10.1038/nature16496.

68. Cardozo Gizzi AM, Cattoni DI, Fiche JB, Espinola SM, Gurgo J, Messina O, *et al.*: **Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms**. *Mol Cell* 2019, **74**:212–222, https://doi.org/10.1016/j.molcel.2019.01.011. e5.

69. • Nguyen HQ, Chattoraj S, Castillo D, Nguyen SC, Nir G, Lioutas A, *et al.*: **3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing**. *Nat Methods* 2020, **17**:822–832, https://doi.org/10.1038/s41592-020-0890-0.
This work introduces the experimental technique for imaging chromosomes called OligoFISSEQ. The technique uses fluorescence in situ sequencing of barcoded Oligopaint probes to visualize targeted genomic regions. The study produces 3D maps of 36 genomic targets across six chromosomes in hundreds to thousands of cells using only four rounds of sequencing.

70. Pierro MD, Potoyan DA, Wolynes PG, Onuchic JN: **Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes**. *Proc Natl Acad Sci USA* 2018, **115**:7753–7758, https://doi.org/10.1073/pnas.1806297115.

71. Fudenberg G, Abdennur N, Imakaev M, Goloborodko A, Mirny LA: **Emerging evidence of chromosome folding by loop extrusion**. *Cold Spring Harbor Symp Quant Biol* 2017, **82**:45–55, https://doi.org/10.1101/sqb.2017.82.034710.

72. Sandholtz SH, MacPherson Q, Spakowitz AJ: **Physical modeling of the heritability and maintenance of epigenetic modifications**. *Proc Natl Acad Sci USA* 2020, **117**:20423–20429, https://doi.org/10.1073/pnas.1920499117.

73. Gürsoy G, Liang J: **Three-dimensional chromosome structures from energy landscape**. *Proc Natl Acad Sci USA* 2016, **113**:11991–11993, https://doi.org/10.1073/pnas.1614535113.

74. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, *et al.*: **OpenMM 7: rapid development of high performance algorithms for molecular dynamics**. *PLoS Comput Biol* 2017, **13**:e1005659, https://doi.org/10.1371/journal.pcbi.1005659.

75. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, *et al.*: **De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds**. *Science* 2017, **356**:92–95, https://doi.org/10.1126/science.aal3327.

76. Contessoto VG, Dudchenko O, Aiden EL, Wolynes PG, Onuchic JN, Pierro MD: **Interphase chromosomes of the Aedes aegypti mosquito are liquid crystalline and can sense mechanical cues**. *bioRxiv* 2022, https://doi.org/10.1101/2022.02.01.478655. 2022.02.01.478655.

77. Zidovska A, Weitz DA, Mitchison TJ: **Micron-scale coherence in interphase chromatin dynamics**. *Proc Natl Acad Sci USA* 2013, **110**:15555–15560, https://doi.org/10.1073/pnas.1220313110.

78. • Su JH, Zheng P, Kinrot SS, Bintu B, Zhuang X: **Genome-Scale imaging of the 3D organization and transcriptional activity of chromatin**. *Cell* 2020, **182**:1641–1659, https://doi.org/10.1016/j.cell.2020.07.032. e26.
This work used massively multiplexed FISH to generate high-resolution snapshots of individual chromosomes, while simultaneously imaging nascent RNA transcripts. This work also used DNA tracing to image the entire human genome while also imaging nascent RNA and nuclear bodies. In particular, this work characterized chromatin domains, compartments, and trans-chromosomal interactions and their relationship to transcription in single cells.

79. • Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, *et al.*: **3D genomics across the tree of life reveals condensin II as a determinant of architecture type**. *Science* 2021, https://doi.org/10.1126/science.abe2218.
This study reports the 3D genome organization for 24 eukaryote species, including animals, fungi, and plants. Hi-C maps and microscopy data show that, in interphase, the nuclei had either telomeres and/or centromeres clustered across chromosomes or were oriented in a polarized state (Type I) or maintaining chromosomal territories within the cell (Type II), a difference attributed to condensin II complex. Experimentally removing condensin II in human cells promotes the formation of centromere clusters but does not affect loop or compartment formation. Theoretically modeling this phenomenon, we have shown how the genome structural ensemble interconverts between these genome architectures - Type I or II for low or high condensin II, respectively.