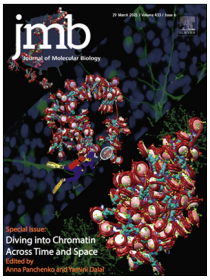
Research Article



A Scalable Computational Approach for Simulating Complexes of Multiple Chromosomes

Antonio B. Oliveira Junior [1,2†], Vinícius G. Contessoto [1,3†], Matheus F. Mello [1,4] and José N. Onuchic [1,2*]

*1 - **Center for Theoretical Biological Physics,** Rice University, Houston, TX, USA*
*2 - **ICTP South American Institute for Fundamental Research,** Instituto de Física Teórica, UNESP – 01140-070, São Paulo, SP, Brazil*
*3 - **Instituto de Biociências,** Letras e Ciências Exatas, UNESP - Univ. Estadual Paulista, Departamento de Física, São José do Rio Preto, SP, Brazil*
*4 - **Chemical Engineering Department,** Military Institute of Engineering, Rio de Janeiro, RJ, Brazil*

Correspondence to José N. Onuchic: Center for Theoretical Biological Physics, Rice University, Houston, TX, USA. *antonio.oliveira@rice.edu (A.B. Oliveira Junior), vinicius.contessoto@rice.edu (V.G. Contessoto), jonuchic@rice.edu (J.N. Onuchic)*
https://doi.org/10.1016/j.jmb.2020.10.034
Edited by Anna Panchenko

Abstract

Significant efforts have been recently made to obtain the three-dimensional (3D) structure of the genome with the goal of understanding how structures may affect gene regulation and expression. Chromosome conformational capture techniques such as Hi-C, have been key in uncovering the quantitative information needed to determine chromatin organization. Complementing these experimental tools, co-polymers theoretical methods are necessary to determine the ensemble of three-dimensional structures associated to the experimental data provided by Hi-C maps. Going beyond just structural information, these theoretical advances also start to provide an understanding of the underlying mechanisms governing genome assembly and function. Recent theoretical work, however, has been focused on single chromosome structures, missing the fact that, in the full nucleus, interactions between chromosomes play a central role in their organization. To overcome this limitation, MiChroM (Minimal Chromatin Model) has been modified to become capable of performing these multi-chromosome simulations. It has been upgraded into a fast and scalable software version, which is able to perform chromosome simulations using GPUs via OpenMM Python API, called Open-MiChroM. To validate the efficiency of this new version, analyses for GM12878 individual autosomes were performed and compared to earlier studies. This validation was followed by multi-chain simulations including the four largest human chromosomes (C1-C4). These simulations demonstrated the full power of this new approach. Comparison to Hi-C data shows that these multiple chromosome interactions are essential for a more accurate agreement with experimental results. Without any changes to the original MiChroM potential, it is now possible to predict experimentally observed inter-chromosome contacts. This scalability of Open-MiChroM allow for more audacious investigations, looking at interactions of multiple chains as well as moving towards higher resolution chromosomes models.

Introduction

Since the human DNA sequence was decoded[1] in early 2000s, significant efforts have been made to understand the three-dimensional (3D) structure of the genome and its role in gene regulation and expression. This challenge gave rise to a plethora of experimental techniques needed to investigate the complexity of chromatin folding.[2–8] Hi-C is central in this effort and it uses a genome-wide high-throughput sequencing technique that provides the contact frequency between all genomic loci pairs. Hi-C data supports the existence of chromosome territories and also identified that the overall genome organization can be described by two major compartments (called A and B) and their respective sub-compartments. Compartment A has a strong presence of genes, thus higher expression, and, as expected, it is more accessible. On the other hand, compartment B is more densely packed and tends to be more located in the core of the chromosomes.[9] Another genome feature identified from these Hi-C maps is the topologically associating domains (TADs), self-interacting partitions in order of mega bases.[10] Additionally, Rao and co-workers identified local domains associated with distinct patterns of histone marks resulting in six sub-compartments; they also observed the presence of loops and observed that loops frequently link promoters and enhancers, indicating their key role in gene activation.[11]

Although Hi-C maps provide information about the genome's organization, the data obtained is an average over a large number cells. It produces a 2D average contact map and not the much desired 3D ensemble of structures. Advances in high-resolution microscopy and polymer modeling, however, shed light on how chromosomes fold in the 3D space.[12–17] Many theoretical models based on polymer physics were designed along the years, which helps in explaining mechanisms such as loops extrusion and phase separation.[18–25] We highlight the Minimal Chromatin Model (MiChroM), which was the first of all these methods and has great prediction power. MiChroM is a theoretical energy landscape model for chromatin folding, which uses the maximum entropy principle.[18,26]

MiChroM energy function was applied to polymer chromosome simulations to generate ensembles of individual human chromosome structures in interphase by using the sub-compartments chromatin types and position of loop anchor loci as inputs. The Hi-C contact maps obtained from the ensemble of chromosome 3D structures reproduced in detail the maps from Hi-C at a 50 kb resolution.[18] Although the model was also used to simulate the combined pair of chromosomes 17 and 18 from human GM12878 lymphoblastoid cells successfully, the computational cost for the simulations increased substantially with the size of the simulation, This has been a limiting factor in simulating large groups of chromosomes. A simulation for the whole human genome, for example, would include more than 150,000 beads at a 50 kb resolution.

To overcome this challenge, we introduce Open-MiChroM, a software tool that implements the MiChroM energy function for chromosome simulation using the OpenMM Python API.[27] Open-MiChroM allows high-performance simulations through Graphics Processing Unit (GPU) acceleration and allows the simulation of more complex and larger systems. Initially, Open-MiChroM is used to generate an ensemble of 3D structures for all individual chromosomes of the cell line GM12878. Using this ensemble of structures, we generate the *in silico* Hi-C maps for each chromosome to validate and determine the efficiency of the new method. This was followed by simulating multiple chromosomes together to demonstrate the scalability power of Open-MiChroM. Initial simulations included chromosomes 1 to 4, generating the *in silico* Hi-C maps for over eight million different structures. The Open-MiChroM code and the trajectories generated for this work are deposited and freely available at the Nucleome Data Bank (https://ndb.rice.edu/). A tutorial page about Open-MiChroM, containing the pipeline to run the chromosome simulations and generate the *in silico* Hi-C maps, is also provided at the NDB website.[28] Comparison to Hi-C data shows that the inclusion of interactions among chromosomes is essential to be able to predict experimental Hi-C maps beyond single chromosome but also reproducing all the experimentally observed contacts between different chromosomes.

Results

Open-MiChroM as tool to simulate human genome

Comparison of Open-MiChroM to previous simulations were performed to validate this new tool and to verify the capabilities and efficiency of this new method. Open-MiChroM simulations reproduced the experimental Hi-C map from the human B-lymphoblastoid cells GM12878[11,18] as shown in Figure 1(A) (upper triangle). The Pearson's correlation between the simulation (lower triangle) and experimental results (upper triangle) is $R = 0.96$ for chromosome 1 at a 50 kb resolution. It is worthwhile to mention that the parameters were obtained from a training set using only chromosome 10 from the GM12878 cell line[18,11] (details about how obtain these parameters are in SI). Chromosome 10 was utilized because of its diversity in patterns of chromosome interactions. We show that these trained parameters are not only transferable to all other chromosomes, as discussed in our earlier work, but also to interactions between chromosomes. This transferability

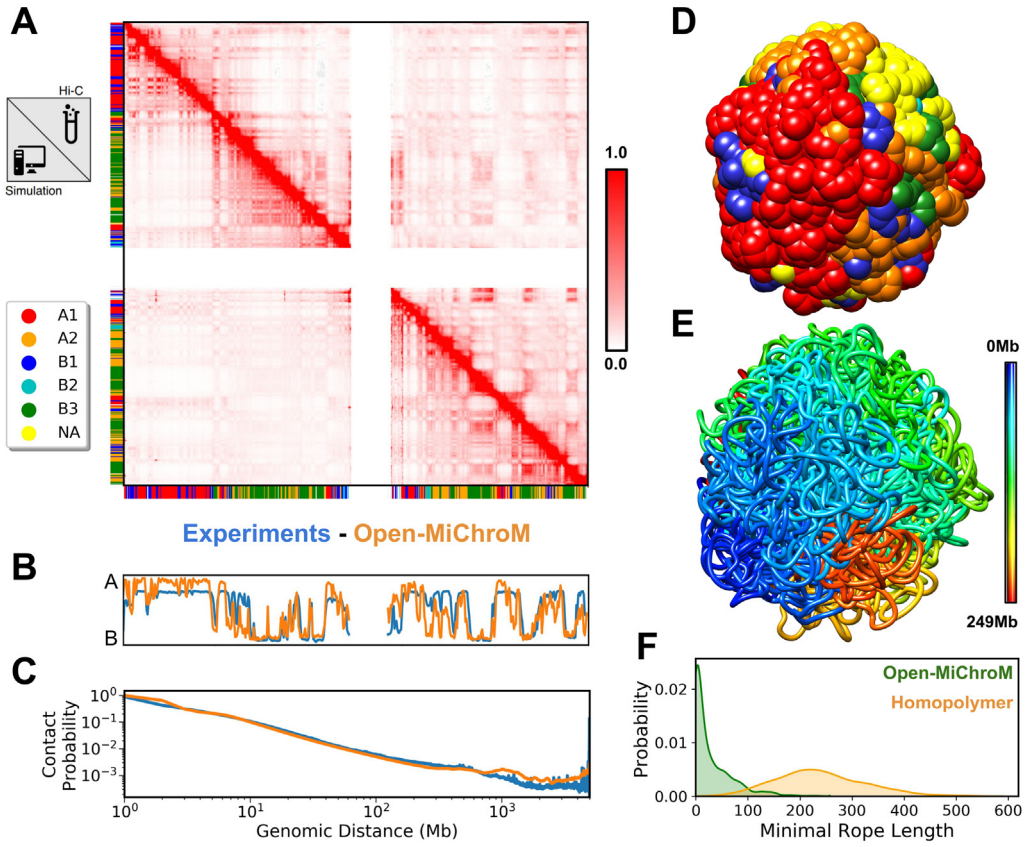


Figure 1. Open-MiChroM simulations of chromosome 1 of the human cell line GM12878. (A) Comparison between the Open-MiChroM *in silico* Hi-C simulations (lower triangle) and experimental Hi-C map (upper triangle).[11] Pearson's correlation between the maps is 0.96. (B) The first eigenvector obtained from the correlation matrix extracted from the Hi-C maps.[9] The orange and blue curves represent experimental data and the Open-MiChroM simulation, respectively. (C) Contact probability as a function of genomic distance. The color scheme is the same as in (B). (D) A representative structure of chromosome 1 highlighting the sub-compartment annotation, following the color scheme shown in (A). (E) The same structure presented in (D) colored by the index locus position. (F) Knot formation distribution for Open-MiChroM and the associated homopolymer.

enables high computing performance by using GPU to run the simulations. As presented here, this hardware acceleration is fundamental for simulating large chromosome complexes such as the human full nucleus to be investigated in future works. Comparison to previous single chromosome simulations[18,11] were performed for all chromosomes in GM12878 and also for additional human cell lines during interphase.[16,28]

Figure 1(B) shows the eigenvectors extracted from the *in silico*, and from the experimental Hi-C results. The agreement between them corroborates the accuracy of Open-MiChroM in determining the compartment annotations A and B. The eigenvector analysis distinguishes the chromatin types A and B: positive and negative values represent the chromatin type A and type B, respectively. Figure 1(C) shows contact probability as a function of the genomic distance. Previous MiChroM simulations exhibited an unprecedented accuracy in reproducing experimental Hi-C maps.[18] Open-MiChroM uses the same MiChroM energy function with one further improvement. Since it is a much more efficient methods, there is no need for including a truncation length in the Ideal Chromosome (IC) term. In the original MiChroM potential, the Ideal Chromosome term was trained with a $d_{cutoff} = 200$ in genome distance and simulated to reproduce the Hi-C maps using a $d_{cutoff} = 500$. Open-MiChroM has an effective way to embed the IC term by using the beads indexation information. Hence, it is now possible to implement the IC term calculation for the entire chain without the need for truncation and, in addition, using less computational resources. The removal of this cutoff leads to an improvement of the polymer scaling when compared to experiments as shown in Figure 1(C). Figure 1(D) shows a representative structure of chromosome 1. The color of the beads is related to the chromatin sub-compartments annotation. Figure 1(E) shows the same chromosome 1 but now the structure is colored by the locus index in the genomic sequence at a 50 kb resolution. Furthermore, Open-

MiChroM has implemented in its software core a function that calculates the knot formation in the chromosome polymer. Reproducing the previous results, Open-MiChroM calculates the Minimal Rope Length[29,30] for each frame of the simulation (details can be found in SI). Figure 1(F) shows the knot distributions for the same chromosome simulated using Open-MiChroM and for a the associated homopolymer. Although it still has a few knots, the chromosome is less entangled than the homopolymer.

The human chromosome simulations follow a procedure described in Methods Section. Open-MiChroM receives a sequence including chromatin types annotation for each chromosome as an input. An initial open configuration of the chromosome chain is built for the minimization step. The structure is collapsed at a high temperature ($T = 5$ in reduced units) for randomizing the initial state for the equilibration and sampling simulations. $T = 1$ was determined in previous work[18] as the temperature that the simulation matches as the temperature during the Hi-C experiments production; the chromosome is folded but still dynamical. After this initial collapse, an annealing procedure is performed, decreasing the temperature from $T = 5$ to $T = 1$. This strategy creates random and uncorrelated configurations for different simulations. The sampling simulation is carried out over 5×10^7 steps via Langevin dynamics with a damping coefficient of 10.0τ, where τ is the time unit. A time step of $\Delta t = 0.01\tau$ was used for all the simulations, storing a structure every 1×10^3 steps. Figure 1(A) (lower triangle) shows the *in silico* Hi-C for chromosome 1 obtained over 40 sampling simulation replicas generating a total of 2×10^5 different conformations.

Open-MiChroM allows simulations of multiple chromosomes chains

The advance of Open-MiChroM makes it possible simulations of multiple chromosome chains, thus allowing us to analyze more than a single chromosome. To determine how multiple chromosome interactions affect the final folding, we have simulated the first 4 human chromosomes of the cell line GM12878. Table 1 describes the number of beads for each polymer chain and the total beads for the multi-chain simulations.

Table 1 Details of the chromosomes set up used in Open-MiChroM multi-chain simulations

	Genome Length (Mb)	Number of Beads (50 kb)
C1	249.50	4990
C2	243.25	4865
C3	198.20	3964
C4	191.40	3828
Total	882.35	17,647

The multi-chain simulations follow the same protocol described in Section "Open-MiChroM as tool to simulate human genome". The structure initialization and the polymer collapse step were performed for each chromosome separately. Once the initial configuration for each chromosome was obtained, the equilibration step starts. The chromosome chains are distributed using the spherical Fibonacci lattices.[31] Each of the 40 replicas has a different initial chromosome distribution. This strategy helps to randomize the neighboring interactions in the chromosome interface. The simulations were carried over 2×10^8 steps and saving a frame every 1×10^3 steps. In total, an ensemble of 8 million chromosomes structures were obtained from 40 replicas. The chromosome trajectories store the XYZ position of each bead of each chromosome at every frame. The data is deposited in the NDB server following the .ndb file format. Open-MiChroM introduces a reading and writing tool for creating a binary version of the .ndb file format, called .cndb. Open-MiChroM also provides a tool for reading the polymer trajectories and generate the *in silico* Hi-C maps and the ensemble of 3D structures.

Figure 2A shows the *in silico* Hi-C map(bottom) from the multi-chain simulation. Open-MiChroM reproduces intra- and inter-chain loci contact observed in the experimental Hi-C maps (upper triangle).[11] The colored squares highlight the intra-chain contacts for the four simulated chromosomes. The Pearson's correlation between the *in silico* and experimental Hi-C is on average $r = 0.94$. Figure 2(B) shows representative structures of the four chromosomes. The color intensity highlights the formation of chromosome territories following the more evident intra-chromosome contacts in Figure 2(A).

Open-MiChroM captures the phase separation between chromatin types in a single and multiple interacting chromosomes

As previously reported, chromatin types A and B tend to phase separate within a chromosome chain.[18,17] This observation is consisted with the need to expose active genes to be expressed. Active genes exist mostly in regions of chromatin type A.[16,19,9] Open-MiChroM has implemented an analysis of the radial density profile for each sub-compartment annotation (The details about RDF calculations are presented in SI). Figure 3A shows the radial distribution for each sub-type (A1, A2, B1, B2, and B3) for a single-chain simulation of chromosome 1 using an ensemble of 2×10^5 structures. As expected, the B sub-type annotations are more often observed in the inner chromosome shell while chromatin types A are mostly present on the chromosome surface. The AB phase separation in the simulations is perturbed when neighboring chromosomes are introduced. The introduction of inter-chain interactions competes with the interac-

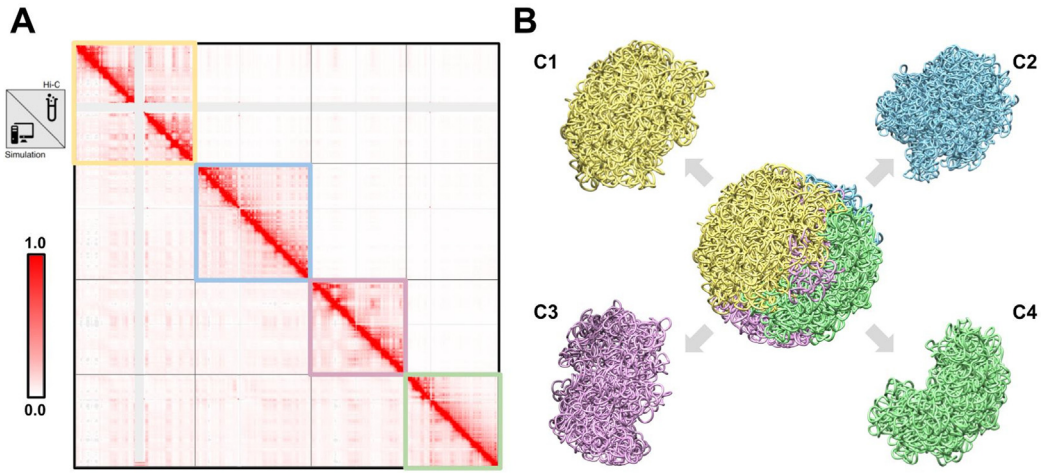


Figure 2. Open-MiChroM multi-chain Hi-C maps and associated representative structures for the combined simulation of the first 4 human chromosomes of the cell line GM12878. (A) Hi-C maps from Open-MiChroM simulations (lower triangle) and experiments (upper triangle). The squared contours highlight the intra-chain contacts for each chromosome. (B) Representative structure of the Open-MiChroM multi-chain simulation containing the chromosomes 1 to 4. The color scheme matches the squared contours used in (A).

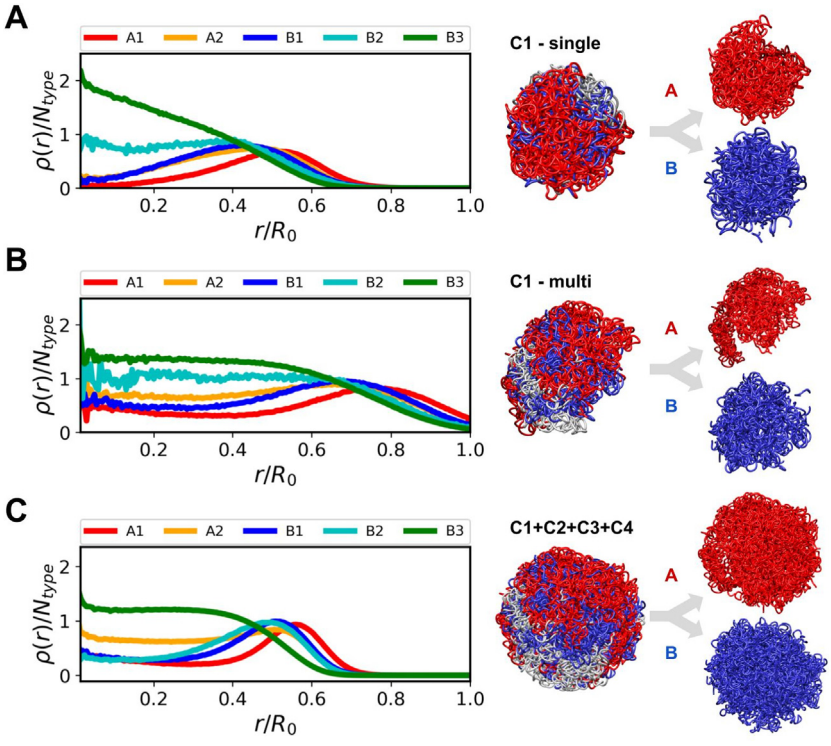


Figure 3. Open-MiChroM single- and multi-chain radial distribution analysis. (A) Open-MiChroM single-chain radial distribution of chromosome 1. The colors represent the different sub-compartment annotations. On the right, it is shown a representative structure highlighting the A-type (red) chromatin located preferentially on the surface and B-type (blue) chromatin buried inside of the chromosome. The gray color represents the centromere region of the chromosome structure. (B) Open-MiChroM radial distribution of chromosomes 1 in a multi-chain simulation. The color scheme is the same as in (A). On the right, it is shown a representative structure of chromosome 1 in the multi-chain simulation.Notice that the structural perturbation due to neighboring interactions is accompanied by a global A-type and B-type phase separation. (C) Open-MiChroM multi-chain radial distribution of the complex of chromosomes 1 to 4. Again, the color scheme is the same as in (A). On the right, it is shown a representative structure of the first four simulated chromosomes. The AB phase separation is observed among different chromosome polymer chains. The radial distribution profile calculation is described in SI.

tions of the single chain. The result is a distortion on the spherical shape of the polymer chain observed in the single-chain simulation since A types tend to move to the surface and B types to the core. Therefore, the chromosome complex maintains independent territories for individual chromosomes and a global AB phase separation, as observed in Figure 3(B) and (C). It is interesting to notice that this distortion in structure for individual chromosomes allow the chromosome complex to satisfy both of these conditions. The liquid-like behavior of chromosomes[32] in interphase allows the distribution chromatin type A on the chromosomes surface, even with the polymer connectivity constraint.

Conclusion

This new theoretical development and computational tool remove existent barriers and allow for extensions of the current simulation beyond individual chromosomes. They represent the first step towards a detailed simulation of the full nucleus genome. In addition, it also allows for future modifications to current models to improve their resolution beyond 50 kb. These advances also make possible to expand our studies to a large number of cell lines. Here we have shown the joint simulation of the four large chromosomes and have shown how inter-chromosome interactions affects the final structure. Although the single chromosome structures still maintain strong similarity to the individual chromosome ones, it is import to highlight that the AB phase separation observed in individual chromosomes is maintained in this large complex.

To achieve these goals, Open-MiChroM is introduced as a software tool to perform dynamical simulations of chromosomes using the Minimal Chromatin Model energy function. Open-MiChroM uses the Python toolkit OpenMM, which enables high computing performance by using GPU to run the simulations. This acceleration is the key for performing simulations of large chromosome complexes and moving towards simulating the full nucleus. There are many other challenges to simulate the human full nucleus, such as determining appropriate initial conditions for the simulations and the effects of interactions with Lamin-A in the nucleus wall, but the theoretical and computational tool towards this goal is now available. Comparison to previous single chromosome simulations[18] were performed to show the power of the method. for example, by using Open-MiChroM, one can now remove the cutoff distance in the Ideal Chromosome function (needed before due to computational limitations) leading to a much improved polymer scaling when compared to experiments. To facilitate result analysis, Open-MiChroM expands the analysis tools in the code core. More specifically, these tools include the knots calculations, radial density distribution, and the *in silico* Hi-C contact map.

Open-MiChroM is also efficient in storing 3D structures of the chromosome polymer chain. The XYZ position of each bead can be exported into different file formats, including the binary format .ndb, named here as .cndb. Python scripts are developed to read and write .cndb files and the conversion of this file format to others commonly used ones. These scripts are made available at the NDB server. Additionally, a tutorial web page, describing, step-by-step usage of Open-MiChroM is included in the NDB server. Expansion of Open-MiChroM for different resolutions will be implemented in the near future. Going beyond human cell lines and tissues, the model is also being expanded to incorporate experimental data from different species and different cell phases. Finally, it is noteworthy that Open-MiChroM is freely available deposited as an MIT license 4, and any contributions to the code are welcome.

Methods

Open-MiChroM is a software tool developed to simulate and analyze the genome organization. Open-MiChroM implements the Minimal Chromatin Model (MiChroM) energy function to simulate chromosome dynamics using the Python API OpenMM software. There are significant advantages in using Open-MiChroM, which are described in Section “Results”. Details about MiChroM, OpenMM, code implementation, and applications are described next.

MiChroM potential

The Minimal Chromatin Model (MiChroM) is a copolymer physics model that uses the maximum entropy principle constrained by three assumptions to incorporate the structural interactions between genomics loci.[20] The energy function takes the following form:

$$U_{MiChroM}(\vec{r}) = U_{HP}(\vec{r}) + \sum_{\substack{k \geqslant l \\ k,l \in \text{Types}}} \alpha_{kl} \sum_{\substack{i \in \{\text{Type}k\} \\ j \in \{\text{Type}l\}}} f(\vec{r}_{ij}) + \chi \sum_{(i,j)\in\{\text{LoopSites}\}} f(\vec{r}_{ij}) + \sum_{d=3}^{d_{cutoff}} \gamma(d) \sum_{i} f(\vec{r}_{i,i+d}), \quad (1)$$

where $U_{HP}(\vec{r})$ is the potential energy of a generic homopolymer with an additional soft-core term to allow for sporadic chain crossing (see SI for details). The following terms result from the constraints applied when implementing the maximum entropy approach. The second term in Eq. (1) corresponds to type-to-type interactions that give rise to the compartmentalization pattern found in Hi-C maps. The third term represents the loop anchor points, to account for the CTCF binding sites. The last term is the translational invariant compaction term, called the ideal chromosome

potential, which assumes that, every time a pair of loci comes into contact, there is a gain/loss of effective free energy, $\gamma(d)$, that is a function of genomic distance d. The spatial range of all interactions is controlled by the function $f(\vec{r}_{ij}) = \frac{1}{2}\{1 + \tanh[\mu(r_c - r_{ij})]\}$, which varies between one at short distances and zero at large separations. All parameters used in this work were obtained from the previous work,[18] see section "GM12878 Hi-C data usage" in SI for more details.

Open-MiChroM implementation and simulation work using OpenMM

OpenMM is a software toolkit for high-performance molecular dynamics simulations. OpenMM allows users to add customized forces with novel energy functions, such as those implemented in Open-MiChroM. OpenMM receives the customize energy function expression as an input and generates an optimized code for high-performance computing. OpenMM supports different hardware architectures, including both CPUs and GPUs.[27]

The Open-MiChroM code uses Python 3 with standard scientific libraries. Figure 4 shows a diagram of the Open-MiChroM simulation pipeline. The code was implemented to accommodate a single or multiple chromosomes chains. Analyses tools and the necessary steps to generate *in silico* Hi-C maps are implemented in the software core. Open-MiChroM initially creates an initial chromosome chain with the chromatin type annotation sequence as an input. Additionally, the loop anchor information is optional and can be provided as a file containing the CTCF loci observed experimentally. The chromatin sequence and the loop anchor points can be provided as text files, or the users can also use the chromosome information provided in the .ndb file. The initial simulation structure can take several forms: a straight line, a random walk structure, a helicoidal shape, or any structure provided by the user in a .ndb file format.

In the case of multiple chains, each chain's initial position is defined using a Fibonacci Lattice distribution. After this initial preparation step, Open-MiChroM is ready to perform simulations

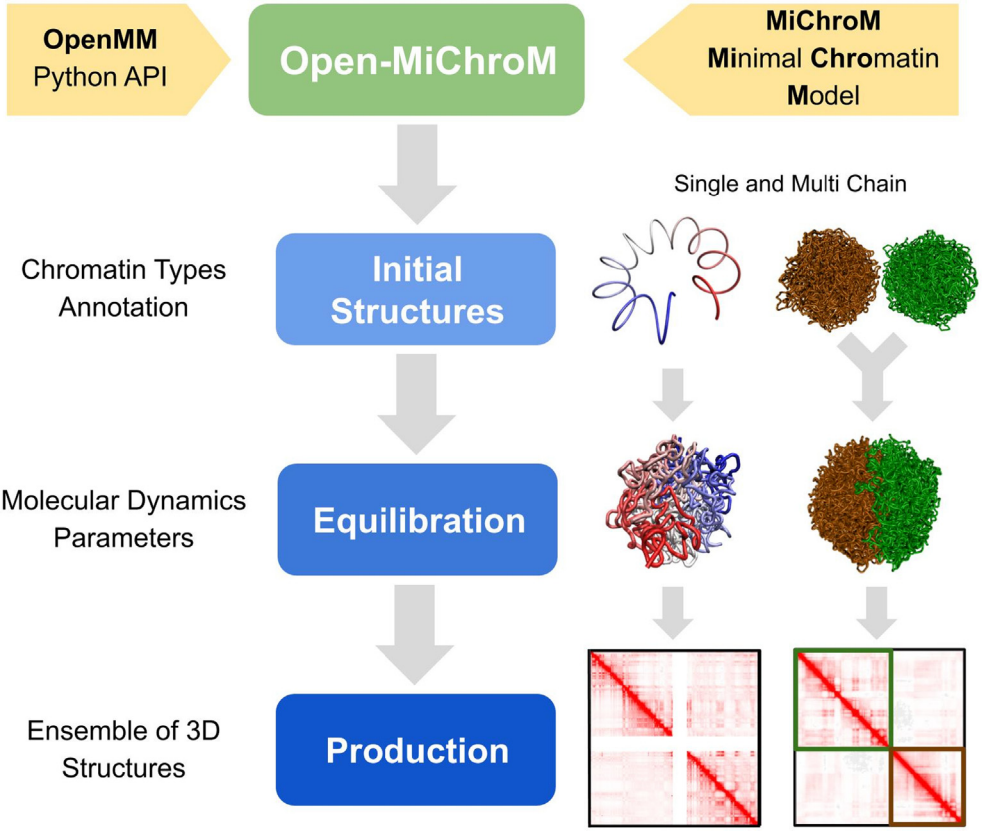


Figure 4. Open-MiChroM implementation diagram. Open-MiChroM combines the Minimal Chromatin Model (MiChroM) with the Python API OpenMM. This implementation allows for chromosome simulations of genomes with high-performance computing by using the hardware acceleration of GPUs. The Open-MiChroM code accommodates simulations of a single or multi chromosome chains. Open-MiChroM reproduces experimental Hi-C maps. The 3D ensemble of chromosome structures is stored in trajectory files, saving them as .ndb (human-readable text file) or as .cndb (binary version of .ndb).

using the energy function described in Section "MiChroM potential". Open-MiChroM equilibration step uses a spherical constraint term for the initial structure collapse. This constraint is optional and is used to speed up the generation of the initial configuration. In the case of multiple chain simulations, equilibration is required until the polymer chains get in contact. After this initial collapse step, this constraint is turned off, and a second equilibration is performed. The second equilibration simulation is a strategy to randomize the initial configuration prior to the sampling simulation. For the simulations presented in this work, the sampling step was carried out over 40 replicas with 5×10^7 steps, storing a frame every 1×10^3 steps that generates a total of two million 3D structures. These numbers were used to produce the in silico Hi-C maps which are compared to experimental ones. The comparison for all human autosomes of the cell line GM12878 are presented in the SI. Trajectory files are stored as .ndb or .cndb files. Analysis parameters such as the potential energies, radius of gyration, knot calculation are outputs when requested with the corresponding flag during the Open-MiChroM simulation. These analyses can also be performed by reading the trajectory files after the simulations.

Open-MiChroM performance analysis in CPUs and GPUs

The Open-MiChroM performance was evaluated based on the simulation time in both CPUs and GPUs. Simulations were performed for chromosome polymer chains at a 50 kb resolution in a range of 1×10^3 to 4×10^5 beads, comparable to sizes of the human chromosome 21 to the full wheat genome. The simulation pipeline follows the scheme presented in Figure 4. A sequence of the chromatin compartment annotation, types A and B, is provided as an input. For benchmarking purposes, the sequences used in this section were randomly generated. The simulations have 2 steps: 1) an initial conformation, either a straight line or a more compact helicoidal structure is chosen. This is followed by an equilibration run needed for compacting the polymer chain and to randomize the initial 3D structure for the next step. 2) Chromosome simulations are performed to sample different conformations which are needed to obtain the 3D structural ensemble of the chromosomes. For this performance analysis, the simulations go over 2×10^5 steps using a time step $\Delta t = 0.01\tau$, where τ is the time unit described in Section "Open-MiChroM as tool to simulate human genome".

Figure 5 shows the performance of Open-MiChroM for different polymer chain lengths. For both CPUs and GPUs, the simulations were averaged over 20 independent runs. The error bars in this benchmarking are negligible. The CPU yellow curve is truncated at the valued of the longest simulation which is shorter than for GPUs (green line). The performance difference between CPUs and GPUs using Open-MiChroM is similar to those obtained in the OpenMM performance analysis in chromosome simulations. The GPU efficiency using Open-MiChroM allows for the investigation of much larger systems. The comparison presented in Figure 5 shows that

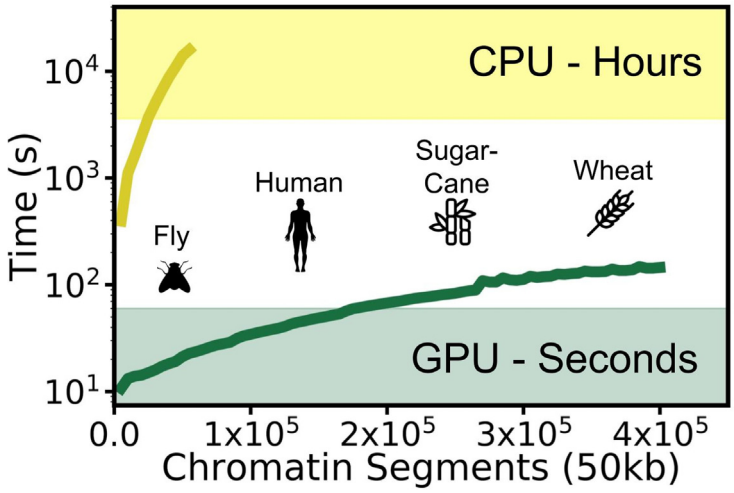


Figure 5. Open-MiChroM performance analyses in CPUs and GPUs. The semi-log plot shows the running time as a function of the polymer size. The times plotted are an average over 20 replicas of 2×10^5 time steps, using a CPU (Intel(R) Xeon(R) Gold 6230 CPU) (yellow) and a GPU (NVIDIA(R) Tesla(R) V100) (green). The shaded regions represent the time scale in seconds (green region) and hours (yellow region). The icons represent the polymer size compared to a full nucleus genome simulation of different organisms.

simulations of a full human genome size with the 46 chromosomes at a 50 kb resolution are now possible using minutes or less of computer time. This also opens the possibility of simulating plant genomes, such as wheat, which have much bigger genomes. It is important to comment that these comparisons only show the ability of Open-MiChroM to scale up in genome size with suitable performance. The results containing the biological comparisons such as between *in silico* and experimental Hi-C maps are presented in Section "Open-MiChroM as tool to simulate human genome".

CRediT authorship contribution statement

Antonio B. Oliveira Junior: Conceptualization, Methodology, Investigation, Data curation, Software, Visualization, Writing - review & editing. **Vinícius G. Contessoto:** Conceptualization, Methodology, Investigation, Data curation, Software, Visualization, Writing - review & editing. **Matheus F. Mello:** Software, Visualization, Writing - review & editing. **José N. Onuchic:** Conceptualization, Methodology, Investigation, Data curation, Project administration, Resources, Funding acquisition, Writing - review & editing.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We want to thank Michele Di Pierro, Ryan Cheng and Peter Wolynes for important discussions and suggestions during the development of this work. JNO wants to thank the hospitality and support from the ICTP/SAIFR and Instituto de Física Teórica, UNESP where much of this research was performed. During this period, JNO wants also to thank the FAPESP (São Paulo State Research Foundation) for the support of his sabbatical. 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. ABOJ was funded by FAPESP (São Paulo State Research Foundation and Higher Education Personnel) Grant 2016/01343-7 and acknowledges the Robert A. Welch Postdoctoral Fellow program. VGC is funded by FAPESP (São Paulo State Research Foundation and Higher Education Personnel), and CAPES (Higher Education Personnel Improvement Coordination) Grants 2016/13998-8 and 2017/09662-7.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jmb.2020.10.034.

Received 26 August 2020;
Accepted 31 October 2020;
Available online 6 November 2020

Keywords:
genome architecture;
Hi-C;
OpenMM;
chromosome simulations

† These authors contributed equally.

References

1. Lander, E.S. et al, (2001). *Nature*, **409**, 860–921. https://doi.org/10.1038/35057062. ISSN 1476-4687.
2. Dekker, J., Rippe, K., Dekker, M., Kleckner, N., (2002). Capturing chromosome conformation. *Science*, **295** (5558), 1306–1311.
3. Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., et al., (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra-and interchromosomal interactions. *Nat. Genet.*, **38** (11), 1341.
4. Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., De Wit, E., Van Steensel, B., De Laat, W., (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nature Genet.*, **38** (11), 1348.
5. Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., et al., (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.*, **16** (10), 1299–1309.
6. Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., et al., (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, **543** (7646), 519.
7. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., et al., (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell*, **174** (3), 744–757.
8. Olivares-Chauvet, P., Mukamel, Z., Lifshitz, A., Schwartzman, O., Elkayam, N.O., Lubling, Y., Deikus, G., Sebra, R.P., et al., (2016). Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature*, **540** (7632), 296.
9. Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.

R., et al., (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326** (5950), 289–293.
10. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., et al., (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485** (7398), 376–380.
11. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E. K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., et al., (2014). A 3D map of the human genome at Kilobase resolution reveals principles of chromatin looping. *Cell*, **159** (7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021. ISSN 0092-8674.
12. Nir, G., Farabella, I., Estrada, C.P., Ebeling, C.G., Beliveau, B.J., Sasaki, H.M., Lee, S.H., Nguyen, S.C., et al., (2018). Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.*, **14** (12), e1007872.
13. Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., et al., (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, **362** (6413), eaau1783.
14. Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., et al., (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, **529** (7586), 418–422. https://doi.org/10.1038/nature16496. ISSN 1476-4687.
15. Di Pierro, M., (2019). Inner workings of gene folding. *Proc. Nat. Acad. Sci.*, **116** (11), 4774–4775.
16. Cheng, R.R., Contessoto, V.G., Lieberman Aiden, E., Wolynes, P.G., Di Pierro, M., Onuchic, J.N., (2020). Exploring chromosomal structural heterogeneity across multiple cell lines. *eLife*, **9**, e60312. https://doi.org/10.7554/eLife.60312. ISSN 2050-084X.
17. Gürsoy, G., Liang, J., (2016). Three-dimensional chromosome structures from energy landscape. *Proc. Nat. Acad. Sci.*, **113** (43), 11991–11993.
18. Di Pierro, M., Zhang, B., Aiden, E.L., Wolynes, P.G., Onuchic, J.N., (2016). Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. U.S.A.*, **113** (43), 12168–12173. https://doi.org/10.1073/pnas.1613607113. ISSN 0027-8424.
19. Di Pierro, M., Cheng, R.R., Aiden, E.L., Wolynes, P.G., Onuchic, J.N., (2017). De novo prediction of human chromosome structures: epigenetic marking patterns encode genome architecture. *Proc. Nat. Acad. Sci.*, **114** (46), 12126–12131.
20. Zhang, B., Wolynes, P.G., (2015). Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **112** (19), 6062–6067. https://doi.org/10.1073/pnas.1506257112. ISSN 0027-8424.
21. Brackley, C.A., Johnson, J., Michieletto, D., Morozov, A.N., Nicodemi, M., Cook, P.R., Marenduzzo, D., (2017). Nonequilibrium chromosome looping via molecular slip links. *Phys. Rev. Lett.*, **119** (13), 138101.
22. MacPherson, Q., Beltran, B., Spakowitz, A.J., (2018). Bottom–up modeling of chromatin segregation due to epigenetic modifications. *Proc. Nat. Acad. Sci.*, **115** (50), 12739–12744.
23. Mirny, L.A., (2011). The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.*, **19** (1), 37–51.
24. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., Mirny, L.A., (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.*, **15** (9), 2038–2049.
25. Krepel, D., Cheng, R.R., Di Pierro, M., Onuchic, J.N., (2018). Deciphering the structure of the condensin protein complex. *Proc. Nat. Acad. Sci.*, **115** (47), 11911–11916.
26. Di Pierro, M., Cheng, R.R., Zhang, B., Onuchic, J.N., Wolynes, P.G., (2019). Learning genomic energy landscapes from experiments. *Model. 3D Conform. Genomes*,, 305.
27. Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y., Beauchamp, K.A., Wang, L.-P., Simmonett, A.C., et al., (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.*, **13** (7), e1005659. https://doi.org/10.1371/journal.pcbi.1005659. ISSN 1553-7358.
28. Contessoto, V.G., Cheng, R.R., Hajitaheri, A., Dodero-Rojas, E., Mello, M.F., Lieberman-Aiden, E., Wolynes, P. G., Di Pierro, M., et al., (2020). The Nucleome Data Bank: web-based resources to simulate and analyze the three-dimensional genome. Nucleic Acids Res., gkaa818., ISSN 1362–4962, doi:10.1093/nar/gkaa818.
29. Lua, R.C., (2012). PyKnot: a PyMOL tool for the discovery and analysis of knots in proteins. *Bioinformatics*, **28** (15), 2069–2071. https://doi.org/10.1093/bioinformatics/bts299. ISSN 1367-4803.
30. Stasiak, A., Katritch, V., Kauffman, L.H., (1998). Ideal Knots |Series on Knots and Everything. Vol. 19, World Scientific Publishing Company, ISBN 978-981-02-3530-7, doi:10.1142/3843.
31. Swinbank, R., Purser, R.J., (2006). Fibonacci grids: a novel approach to global modelling. *Q.J.R. Meteorolog. Soc.*, **132** (619), 1769–1793. https://doi.org/10.1256/qj.05.227. ISSN 0035-9009.
32. Maeshima, K., Ide, S., Hibino, K., Sasai, M., (2016). Liquid-like behavior of chromatin. *Curr. Opin. Genet. Develop.*, **37**, 36–45.