Computational and Structural Biotechnology Journal 23 (2024) 3537–3548

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Method Article

Multiscale molecular modeling of chromatin with MultiMM: From nucleosomes to the whole genome

Sevastianos Korsak [a,b], Krzysztof Banecki [a,b], Dariusz Plewczynski [a,b,*]

[a] Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland
[b] Centre of New Technologies, University of Warsaw, Warsaw, Poland

ARTICLE INFO

Dataset link: GSE63525

Dataset link: https://github.com/CSOgroup/CALDER2

Dataset link: ENCSR637XSC

ABSTRACT

Motivation: We present a user-friendly 3D chromatin simulation model for the human genome based on OpenMM, addressing the challenges posed by existing models with use-specific implementations. Our approach employs a multi-scale energy minimization strategy, capturing chromatin's hierarchical structure. Initiating with a Hilbert curve-based structure, users can input files specifying nucleosome positioning, loops, compartments, or subcompartments.

Results: The model utilizes an energy minimization approach with a large choice of numerical integrators, providing the entire genome's structure within minutes. Output files include the generated structures for each chromosome, offering a versatile and accessible tool for chromatin simulation in bioinformatics studies. Furthermore, MultiMM is capable of producing nucleosome-resolution structures by making simplistic geometric assumptions about the structure and the density of nucleosomes on the DNA.

Code availability: Open-source software and the manual are freely available on https://github.com/SFGLab/MultiMM or via pip https://pypi.org/project/MultiMM/.

1. Introduction

Chromatin, which exhibits a multi-scale organization [1–4], is composed of nucleosomes as its fundamental units ($1; kb - 10; kb$). Each nucleosome consists of eight histone proteins that interact with DNA primarily through electrostatic forces [5–8]. These nucleosomes, often referred to as the "quantum" of chromatin, play a key role in the compaction of DNA, which wraps around the histone core approximately 1.6 times, forming the classic "beads-on-a-string" structure. This biophysical organization is essential for regulating DNA accessibility, contributing to the dynamic nature of chromatin's role in gene expression, replication, and repair. At a higher scale ($50\ kb - 200\ kb$) [9–11], loops form, driven by Smc complexes with ring-like topology acting as loop extruders and CTCF proteins serving as extrusion barriers [12]. Large blocks of chromatin can be classified as A and B compartments ($\sim 2\ Mb$), and they imply a phase separation which is highly correlated with differential densities of loops and nucleosomes, a phenomenon extendable to four or five subcompartments [11,13,14]. It is worth noting that compartmentalization is highly correlated with epigenetic states and histone modifications [15,16] and can influence gene expression [17]. Interestingly, recent high-resolution Hi-C data has provided scientists with the opportunity to detect compartments at kilobase resolution as well [14]. Further complexity arises from the interaction of chromosomes ($50\ Mb - 250\ Mb$), each predominantly interacting with itself, within the genome [18]. Notably, the nuclear lamina introduces an additional layer of intricacy, as specific chromatin regions interact with it [19,20]. Consequently, the hierarchical structure of chromatin implies different biophysical laws applicable at particular scales.

The resolution-specific impact of experimental data affects the complexity of chromatin analysis. Experiments like MNase-Seq or ATAC-Seq can determine the positioning of nucleosomes [21–23], and they can make a distinction between open and closed chromatin [24]. Nucleosome prositioning is not a trivial problem and special tools have been implemented for this task [24–27]. Loop determination relies on 3C-type experiments like Hi-C, ChIA-PET, or Hi-ChIP [13,14,28,29], which give us information about the genome-wide interactions. At broader scales, compartmentalization patterns are revealed by analyzing the sign of the first eigenvector in Hi-C data [13,14], while more advanced models like Calder can classify or cluster chromosomes into subcompartments [30]. Chromosomal territories can be easily observed from 3C-type experiments, which reflect a limited amount of intra-chromosomal interactions.

* Corresponding author.
E-mail address: Dariusz.Plewczynski@pw.edu.pl (D. Plewczynski).

https://doi.org/10.1016/j.csbj.2024.09.025
Received 9 August 2024; Received in revised form 28 September 2024; Accepted 29 September 2024
Available online 2 October 2024

Modeling various scales of chromatin structure poses a formidable challenge due to the enigmatic nature of biophysical forces within chromatin. Nucleosomes, with a relatively higher degree of understanding, have been extensively studied through simulations capturing histone interactions with DNA mediated by electrostatic forces [5–8,31]. However, integrating these models for large chromatin regions is computationally impractical, and this is the reason why geometrical approaches for the simulation of nucleosomes have been developed [8,32]. Stochastic models at the loop scale incorporate a rebinding mixing probability for Smc complexes diffusing in one dimension, hindered by CTCF barriers [10,12,33–36]. In genome-wide models, fixed harmonic bond forces [37] are often assumed to circumvent the computational complexities of Monte-Carlo simulations, with equilibrium lengths and spring coefficients fine-tuned based on loop strength or PET-count from Hi-C, Hi-ChIP, or ChIA-PET data. Compartmentalization forces are frequently represented using a block-copolymer model [38], simulating long-range attractive interactions between compartments sharing the same compartment. Exploring interactions with the nuclear lamina [19,39], such as the attraction of the B compartment near the lamina [40,41] or the increased density of small chromosomes near the nucleus, provides insights into more realistic structures.

Chromatin modeling is a vibrant field with numerous publications each year, reflecting a wide array of modeling approaches. A significant subset of chromatin modeling focuses on optimization-based methods, wherein models are refined to best fit experimental contact data from 3C-type assays [42–46]. These optimization procedures not only enhance model accuracy but also serve as a form of validation. However, such models are often primarily mathematical in nature and do not contribute substantially to the biophysical understanding of chromatin processes. Moreover, they typically apply uniform rules across all chromatin scales, which lacks biophysical validity. Conversely, some models concentrate on specific chromatin scales, such as stochastic loop extrusion models [10,33–36,47–49]. These models consider the stochastic binding and unbinding of loop extrusion factors (e.g., cohesin and condensin) and their interactions with boundary elements like CTCF proteins. While these stochastic models can reconstruct experimental heatmaps through ensembles of structures, they are limited by their focus on TAD scales and the computational demands of Monte Carlo simulations, which hinder integration with other biophysical processes, such as compartmentalization. Furthermore, nucleosome simulations [5,6,8,31], which mainly target very high resolutions, but they are difficult to be generalized for higher scales. Another category encompasses models based on biophysical force-fields that operate from random initial structures [31,37,39,50–53]. These models often utilize data-driven forces that yield biophysically reasonable structures, meaning that they are optimizing energy functions derived from experimental observations. Some of these models are made to be applied for specific kind of data like single cell Hi-C. Although they can efficiently model data across different chromatin scales, they often rely on phenomenological approximations, lacking deeper biophysical insights. Furthermore, they struggle to effectively represent all chromatin scales, from nucleosomes to chromosomal territories and lamina interactions. Benchmarking and comparing these models is complex because each focuses on modeling different scales or biophysical aspects of chromatin. They employ varied validation methods, which are often insufficient due to the limited availability of experimental data. Additionally, many models lack long-term support, are poorly documented, and are difficult to use.

In this context, MultiMM emerges as a novel human-specific model classified within the latter category, aiming to concurrently address all chromatin scales. To achieve this, we employ cost-efficient strategies, such as using Hilbert curves for initial structures and geometrical interpolation for nucleosomes. Despite the challenges in chromatin model validation, we offer several default working parameters to enhance the convenience of our tool. Furthermore, MultiMM aims to provide a user-friendly, open-source software platform that allows users to experiment with force parameters, configurations and data. Notably, MultiMM can optimize 3D chromatin structures in minutes with minimal input data and has been tested across various platforms and datasets, including both population-averaged and single-cell 3C-type experiments. Our model can be seen as a continuation of our previous chromatin "Spring model" [37], made to run for much larger structures taking into account the biophysics of each one of the scales.

2. Materials and methods

2.1. The general scheme of the model

In its initial steps, MultiMM adapts the provided loop data to match the simulation's granularity, downgrading the data accordingly. Compartmentalization data are imported from Calder [30] or any other software that can provide information about compartmentalization, and they are converted into a discrete value vector, where positive values represent A compartment and negative B compartment. In the case of nucleosomes, we assume that their density is inversely proportional to the ATAC-Seq signal, where a user-determined maximum amount of nucleosomes is assigned to each bead.

MultiMM constructs an initial structure based on a Hilbert curve after importing user-provided input data. The Hilbert curve, originating from the study of fractals [54], represents a structure characterized by self-similarity, a property commonly used to describe natural geometries [55–57]. The central concept of a fractal involves the iterative expansion of a one-dimensional line pattern into higher dimensions, usually non-integer fractal dimensions, resulting in highly folded structures that avoid self-intersection. The self-avoiding nature and inherent compaction of this initial structure provide significant advantages, notably expediting the energy minimization process. Each chromosome is represented by a distinct polymer chain, with interactions between different chromosomal chains being permitted. The initial positions of chromosomes on the Hilbert curve can be randomized by shuffling their starting and ending points within the initial structure.

Subsequently, a multi-scale molecular force-field is employed, encompassing strong harmonic bond and angle forces between adjacent beads, along with harmonic spring forces of variable strength to model the imported long-range loops. For compartment modeling, a long-range block-copolymer potential with discrete energy levels is applied, where stronger forces denote robust attractive interactions between B compartments. The number of energy levels corresponds to the count of distinct (sub)compartments. Intra-chromosomal interactions are also modeled as long-range harmonic bond forces, with the user responsible for providing relevant interactions to influence the resultant structure. The genome is folded within a spherical container with radii $R_2 = \lambda R_1$, $\lambda > 1$, where the smaller radius represents a boundary condition for the nucleolus. Addressing lamina interactions is challenging due to limited access to publicly available Lamin B1 ChIP-seq data [58,59]. Consequently, we model these interactions by assuming that the B compartment is primarily attracted to the lamina, where $r = R_1$ or $r = R_2$. To represent this, we utilize a (sub)compartment-specific trigonometric potential with minima near the spherical walls for the B compartment and in the intermediate region for the A compartment. Additionally, acknowledging that smaller chromosomes tend to be closer to the nucleolus than larger ones, we incorporate an attractive potential proportional to the chromosomal size and inversely proportional to the distance from the center. By minimizing the energy of this force-field by using an OpenMM integrator of preference, we can conclude to genome-wide structures.

For the implementation of the model, the python front-end of OpenMM was used [60]. The energy minimization can be accelerated across multiple CPUs, or with platforms CUDA and OpenCL. Appropriate force-field parameters are selected by default, though the users may change them according their special preferences.

Following the execution of MultiMM, a file containing the minimized genome-wide structure is generated in `.cif` format. This structure can

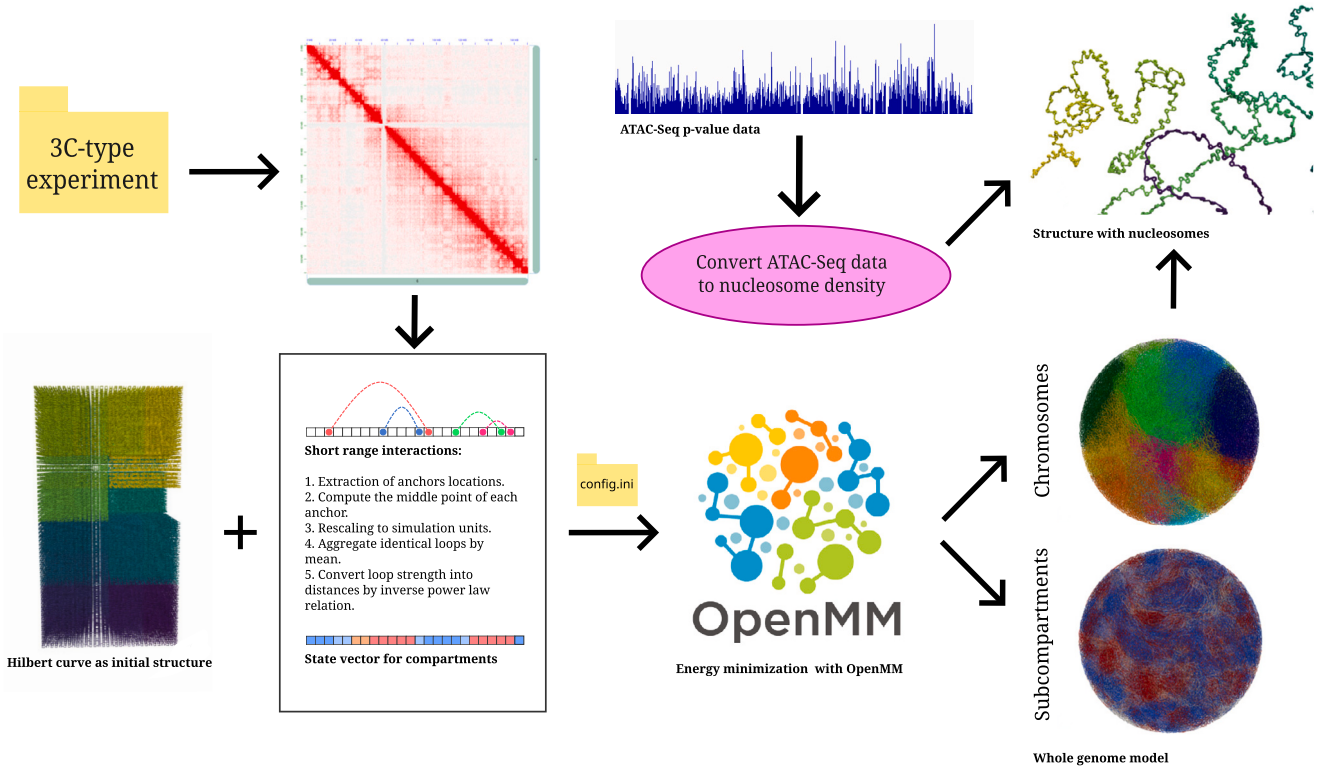


Fig. 1. Flowchart of MultiMM. The process begins with the user selecting a 3C-type experiment and extracting loop and (sub)compartment interactions. MultiMM then takes these data as input and converts them into simulation quantities through unit rescaling. These quantities are imported into the OpenMM force field and applied to an initially generated Hilbert curve structure. After a period of simulation, the final genome-wide structures are produced. If ATAC-Seq data is provided, MultiMM can also generate structures inclusive of nucleosomes.

be conveniently visualized using the pyvista library [61], with accompanying data for chromosome and (sub)compartment coloring readily available. Furthermore, distinct `.cif` structures for each chromosome are extracted, facilitating focused examination of individual chromosomal configurations.

2.2. Input data and pre-processing

The minimal set of data that is needed to make our software to work, is a 3C-type like Hi-C, ChIA-PET, Hi-ChIP, which can be averaged or single cell. The user needs to prepare data and provide locations of interactions across the genome. Usually population averaged data are modeled by tuning the distances as an inverse power law of the interaction strength, whereas single cell data are modeled as binary interactions.

Firstly, MultiMM imports loop data. The only restriction for these data, is to be in the `.bedpe` format. The first three columns should determine the chromosome and genomic coordinates of the left anchor, whereas the next three the coordinates of the right anchor. The seventh column is the strength of the loop $S_{i,j}$, where i and j refer to the anchor locations. Upon importing the loops, MultiMM calculates the midpoint of each anchor, converts these midpoints into genomic coordinates using min-max integer normalization, and aggregates the strength of loops with identical indices by imputing their values with the average (Fig. 1). Then there are two options of modeling, the simplest one is to model all loops with fixed equilibrium lengths[1] $d_{i,j} = 0.1\,u$ or alternatively to correlate the equilibrium length as inversely proportional to the loop count $d_{i,j} \sim 1/S_{i,j}^{2/3}$. The equilibrium lengths of the loops are normalized to be within the interval of $0.1\,u < d_{i,j} < 0.2\,u$. An exception are single cell data, where all the distances should always be fixed. Loop data are essential to run the software, without them it returns an error.

Apart from looping data, MultiMM is capable of importing compartmentalization data as well. The user needs to provide a `bed` file with the coordinates of all regions and their (sub)compartment label. The `bed` file needs to have four columns with the chromosome of interest, the start, end of the region and the compartment label in the third column (A and B). Alternatively, the user can provide subcompartments as input as well (A1, A2, B1, B2). Having imported (sub)compartments, MultiMM generates an array with epigenetic spins s which takes values $s \in \{-1, 0, 1\}$ in case of compartments and $s \in \{-2, -1, 0, 1, 2\}$ in case of subcompartments, with the positive values to correspond to A compartments. Notably, we assume that in case that the spin $s = 0$, it means that it is undefined due to the lack of data, and thus there is not force for this region. Useful software that can be used for the generation of subcompartments from Hi-C data is Calder [30].

2.3. Benchmarking

As previously stated, MultiMM is capable of modeling chromatin structures within minutes. However, computational time is inherently dependent on the available resources and the level of granularity required for the simulation. To prevent structural collapse due to high interaction density, larger structures are generally preferable. For instance, modeling 10^4 loop interactions with $N_{beads} = 10^3$ would result in collapse, necessitating the use of higher granularities such as $N_{beads} = 10^4$ or $N_{beads} = 10^5$. Further increases in granularity contribute only to computational overhead without yielding significant structural changes.

We evaluated computational time on both CPU and GPU using two systems: a Lenovo ThinkPad (Fig. 2a) with an Intel i7 processor and NVIDIA GPU, and a Lenovo Hopper server (Fig. 2b). Our findings in-

[1] Distance is measured in simulation units (u) instead of OpenMM nanometers (nm) to distinguish simulation units from real physical distances due to the coarse-grained nature of the simulations.

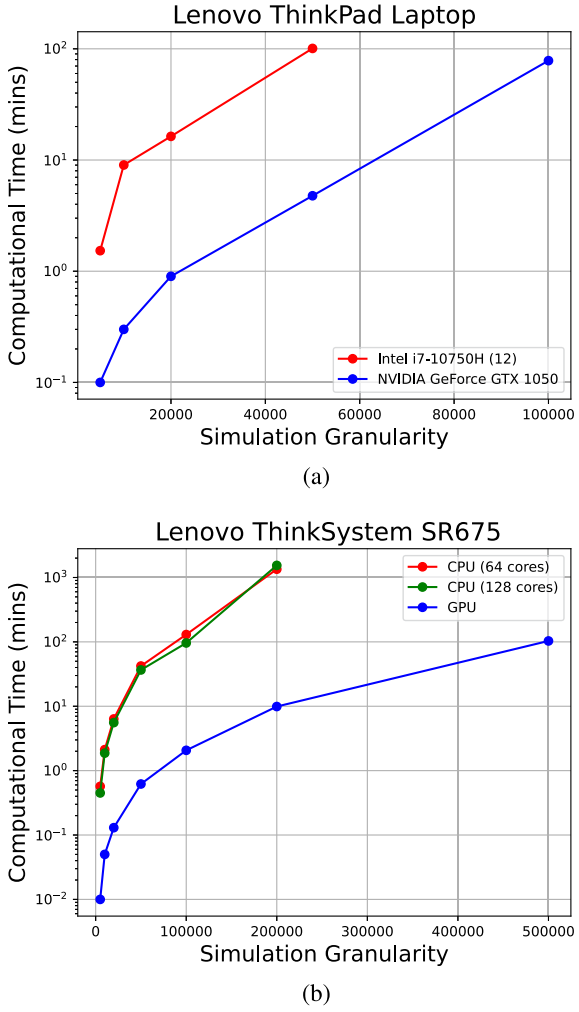


Fig. 2. Benchmarking of MultiMM based on computational time: (a) performed on a Lenovo ThinkPad laptop and (b) on a Lenovo Hopper server. The y-axis represents computational time in minutes (logarithmic scale), while the x-axis corresponds to varying values of simulation granularity, N_{beads}. In subfigure (b), the server allows simulations of one order of magnitude larger than those achievable on the ThinkPad.

dicate that it is feasible to model structures with a granularity of up to $N_{beads} = 5 \times 10^4$ on a home laptop within about 8 minutes. Larger structures with $N_{beads} = 10^5$ can be simulated in a home laptop with GPU within 1 hour and 18 minutes. While simulations using 12 CPU cores are possible, they are significantly slower, taking several hours for larger structures. Nonetheless, we consider this granularity sufficient for most applications, as it adequately captures the loop interactions observed in 3C-type experiments. On the Lenovo server, we achieved much faster results, successfully simulating granularity up to $N_{beads} = 5 \times 10^5$ on a GPU within two hours, producing the detailed images presented in this paper. Tests with 64 and 128 CPU cores on the same server demonstrated that structures with similar amount of monomers can still be modeled within a day.

2.4. Initialization of the simulation

To set up the simulation, the user needs to provide a configuration file. Within the configuration file, the user needs to define the datasets that he uses, and the forces that are enabled. Not all forces are always needed. MultiMM can either model specific regions of the data, or the whole genome. In case, the user wants to model the whole genome, they need to enable additional forces that are related to interactions with lamina, compartments and chromosomes. In Table 1 we can see a summary of the forces that are enabled by default and the forces that are needed in genome-wide simulation. In this manner, forces that are connected to the spherical container are essential to reproduce the final structure. On the other hand, there are some optional forces that can be enabled only if the user provides appropriate data. For instance, block-copolymer forces cannot be modeled without data that provide information about the compartments, or nucleosome interpolation cannot run without data that provide information about the nucleosome density (i.e. ATAC-Seq or MNase-Seq).

Table 1
In this table, the names of all forces in simulations are shown. The "Default" column indicates whether the particular force field is enabled by default. The other two columns suggest whether the user should enable the particular force-field for modeling a specific region or the whole genome.

Force-field Name	Default	Specific Region	Genome-wide
Backbone harmonic bonds	Enabled	Yes	Yes
Harmonic angle forces	Enabled	Yes	Yes
Loop forces	Enabled	Yes	Yes
Excluded volume (LJ)	Enabled	Yes	Yes
Container force	Disabled	No	Yes
Chromosomal blocks	Disabled	No	Optional
Compartment blocks	Disabled	Optional	Optional
Subcompartment blocks	Disabled	Optional	Optional
Interaction with lamina	Disabled	No	Optional
Attraction of nucleolus	Disabled	No	Yes
Nucleosome interpolation	Disabled	Optional	Optional

Table 2
Table of default simulation parameters. Values inside brackets represent intervals of values.

Parameter Name	Symbol	Value	Units
Harmonic Bond Strength	k_b	3×10^5	$kJ/(mol \cdot u^2)$
Harmonic Bond Equilibrium Distance	ℓ	0.1	su
Harmonic Loop Strength	k_l	3×10^4	$kJ/(mol \cdot u^2)$
Harmonic Loop Equilibrium Distance	d	[0.1, 0.2]	u
Harmonic Angle Strength	k_s	10^2	$kJ/(mol)$
Harmonic Equilibrium Angle	ϑ	π	radians
LJ Strength	ϵ	10^2	kJ/mol
Length Scale of LJ	σ	0.05	u
LJ Power	α	3	–
Container Strength	C	10^3	$kJ/(mol \cdot u^2)$
Lamina Attraction Strength	B	4×10^2	kJ/mol
Attraction of the Nucleolus	G	10^2	kJ/mol
Strength of Chromosomal Attraction	E_n	10^{-5}	kJ/mol
Chromosomal Scale Factor	k_C	0.3	u^{-4}
Compartment Energy Levels	E_c	[1, 2]	kJ/mol
Maximum Nucleosomes per Bead	n_{max}	4	–
Nucleosome Radius	r_n	10^{-2}	u
Points per Nucleosome	p_n	20	–
Zigzag Angle	ϕ_{norm}	$\pi/5$	radians

The default parameters of the force-field are provided in Table 2. These parameters are tested to work well with the genome-wide simulation. Users are encouraged to experiment with parameters so as to receive results that they agree with their biophysical intuition.

The simulation was tested on Linux Debian, Ubuntu and Red-Hat based operating systems, with CPU, CUDA and OpenCL parallelization. We suggest users run it in Linux environments. It is possible to run MultiMM in Windows and Macintosh as well, though users may expect some issues. For example, libraries like pybigwig that is used for loading data for nucleosome interpolation, do not work with Windows. Similarly, although it is possible to run MultiMM in Macintosh computers, their technology lacks of CUDA parallelization, and it may result in longer computational times.

2.5. Force-field

To present the force-field of the simulation we must take into consideration that it can be factored in four basic terms,

$$E = E_{\text{pol}} + E_{\text{lamina}} + \sum_{i,j\in\text{loops}} k_{i,j}(r_{i,j}-d)^2 + \sum_{c\in\{A,B\}} E_c \exp\left(-\frac{r^2}{2r_0^2}\right) \tag{1}$$

where E_{pol} represents terms that are responsible for the formation of the main structure of the polymer, E_{lamin} is the interactions with lamina, the third term represents loop interactions, and the fourth one block-copolymer interactions [38]. In this force-field there are many different parameters. In the following text, we describe the default values of them, however, the user is able to easily change them by configuring the input configuration file.

2.5.1. Polymer interactions

The polymer interactions term includes basic interactions that are responsible for the formation of the polymer structure. Therefore, it can be described by the following terms,

$$E_{\text{pol}} = \sum_i k_b(r_{i,i+1}-\ell)^2 + \sum_i k_s(\theta_{i,i+1}-\vartheta)^2 + \epsilon\left(\frac{\sigma}{r}\right)^{\alpha} \tag{2}$$

the first term fixes the distances between adjacent beads. We assume that the default value of the equilibrium distance is $\ell = 0.1$ u, whereas the spring strength is $k_b = 3\times10^5$ kJ/(mol · u^2). The second term is an angle bond force responsible for the polymer stiffness. Therefore, we assume that the equilibrium angle is $\vartheta = \pi$ and therefore tends to make the polymer linear. The strength of the angle force is soft $k_s = 100$ kJ/mol. The third term represents a repelling Lennard-Jones potential. Usually, we assume that $\epsilon = 100$ kJ/mol and $\alpha = 3$.

2.5.2. Looping interactions

We described the loop potential as follows,

$$E_{\text{loop}} = \sum_{i,j\in\text{loops}} k_{i,j}(r_{i,j}-d)^2 \tag{3}$$

where i and j are locations where we know from the experiment that they are in close proximity due to some loop force. At least loops from experiment are needed to run MultiMM. The loops should be in the format of a `.bedpe` file. The equilibrium length $d_{i,j}$ is inversely proportional to the loop count $d_{i,j} \sim 1/S_{i,j}^{2/3}$. The equilibrium lengths of the loops are normalized to be within the interval of $0.1\,u < d_{i,j} < 0.2\,u$. The harmonic bond strength k_l is always fixed $3\times10^4\ kJ/(mol\cdot u^2)$. In case of single cell data modeling, where the interactions are binary and the interaction strength takes values $S_{i,j} \in \{0,1\}$, all distances are fixed and equal to $0.1\,u$.

2.5.3. Block-copolymer interactions

We described compartment forces with the block copolymer potential [1,38,62]

$$E_{\text{block}} = \sum_{c\in\{A,B\}} E_c \exp\left(-\frac{r^2}{2r_0^2}\right) \tag{4}$$

where we assume that r_0 is the proximity of the long range compartmentalization interaction, and E_c are the energy states for compartments c. Depending on the data that we have, we may have compartments (i.e. A, B) or subcompartments (i.e. $A1$, $A2$, $B1$, $B2$). In case that we have compartments and $s \in \{-1,0,+1\}$ then we can write

$$E_c = E_a\delta(s_1-1)\delta(s_2-1) + E_b\delta(s_1+1)\delta(s_2+1) \tag{5}$$

whereas in case that we have subcompartments $s \in \{-2,-1,0,1,2\}$ then we can generalize energy states to four,

$$\begin{aligned} E_c = {} & E_{a1}\delta(s_1-2)\delta(s_2-2) + E_{a2}\delta(s_1-1)\delta(s_2-1) \\ & + E_{b1}\delta(s_1+1)\delta(s_2+1) + E_{b2}\delta(s_1+2)\delta(s_2+2), \end{aligned} \tag{6}$$

where the energy levels are set so as to have stronger attraction for B compartments which are linked to more condensed regions. Therefore, we can set the parameters $E_a = -1$ kJ/mol and $E_b = -2$ kJ/mol for compartments or $E_{a1} = -1$ kJ/mol, $E_{a2} = -1.33$ kJ/mol, $E_{b1} = -1.66$ kJ/mol and $E_{b2} = -2$ kJ/mol in case that we have subcompartments. The default proximity parameter for compartmentalization is $r_0 = (R_2 - R_1)/20$.

In case there are not enough loop interactions (eq. (3)) in our dataset and most of the loops have small lengths or compartmentalization data are missing, we can use a force to make chromosomes more globular. This force does not have a direct biological meaning and this is why it is disabled by default. Therefore, there is another one block-copolymer kind of potential that operates over chromosomes and it is disabled by default. This potential can be described by the equation,

$$E_{\text{chrom}} = \sum E_n(k_C r^4 - br^3 + cr^2) \tag{7}$$

where $k_C = 0.3\,u^{-4}$ is a parameter that controls how flat the potential is near the minimum, and $E_n = 10^{-5}$ kJ/mol. By controlling the parameter E_n we can have very loose chromosome when it is close to 0 and very globular when it is close to 1. We suggest modeling with small values of the energy, since chromosomes are considered loose. Parameters $b = 1\,u^{-3}$ and $c = 1\,u^{-2}$ exist in equations only for the complete definition of units.

2.5.4. Lamina interactions

In general case, we assume that the whole genome consists of by separate polymer chains that represent chromosomes. The whole chromatin structure is enclosed within a spherical container with radius R_2 that represents the wall of nucleus. Furthermore, we assume that there is an empty space with radius R_1 in the center of the polymer structure that represents nucleolus. In this perspective, we have two walls and polymer structure lives between them.

Apart from the spherical boundary walls, we have two other types of interactions with lamina: attraction of B compartment to lamina, and attraction of smaller chromosomes to the nucleolus (Fig. 3b). Consequently, this potential has the following form,

$$E_{\text{lamina}} = E_{\text{bc}} + E_{\text{Bl}} + E_G \tag{8}$$

where the first term represents the lamina boundary condition in the walls R_1 and R_2,

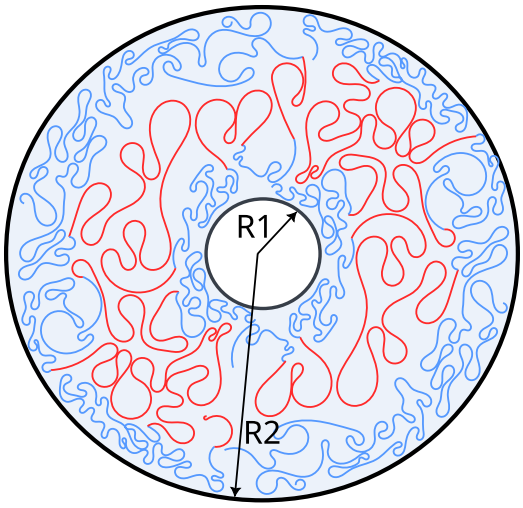
$$E_{\text{bc}} = C\left(\max(0, r-R_2)^2 + \max(0, R_1-r)^2\right) \tag{9}$$

where C is the strength of the wall and it is given value $C = 1000$ kJ/(mol · u^2) by default. The radii R_1 and R_2 can be correlated with the simulation length L so as to be $R_1 = (L/(5\times10^4))^{1/3}$ $[u]$ and $R_2 = 3.5R_1$ (Fig. 3). Therefore, R_2 represent the outer boundary of the nucleus lamina, whereas R_1 represents the inner nucleolus [63,64].

The second term represents the attraction of B compartment with lamina [40,65–67] as it is shown in Fig. 3a. This term becomes minimum near the walls R_1 and R_2, and thus it has the following functional form,

$$E_{Bl} = B\left(\sin^8\left(\frac{r-R_1}{R_2-R_1}\right) - 1\right)(\delta(s+1) + \delta(s+2)) \tag{10}$$

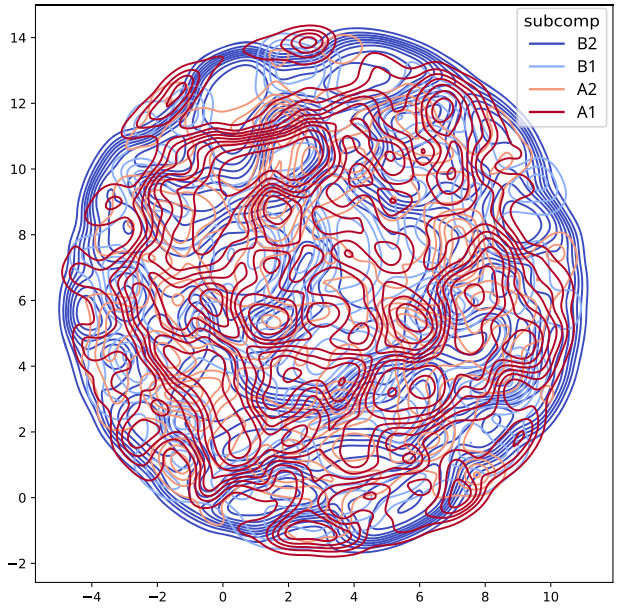
where $\delta(\cdot)$ represents delta Kronecker function which becomes 1 only when the compartment index s is negative and therefore the region is within the B compartment. We set B to be by default 400 kJ/mol representing a soft boundary effect. The effect of the lamina interactions in the final structure can be seen from Fig. 4 where we have plotted the distribution of subcompartments in space. In reality, it is not so easy to have a perfect phase separation in B compartments near lamina as it is presented in Fig. 3a because in many situations A compartment segments are in close genomic distance with B compartment segments,



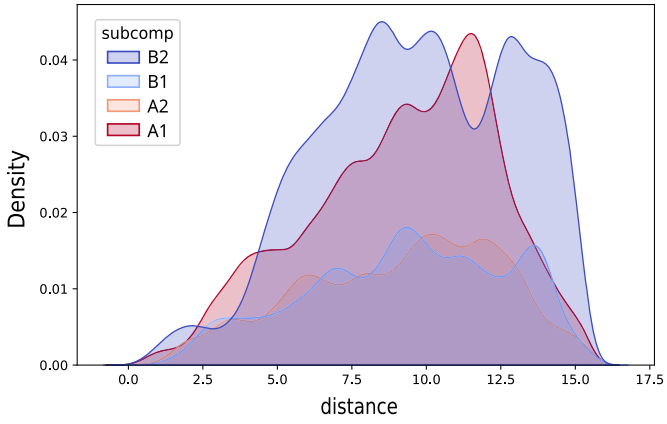
(a) Compartment distribution.

(b) Chromosome distribution.

Fig. 3. The location of compartments and chromosomes in the genome. (a) Compartments are localized in such a way that they are attracted by the outer and inner lamina. (b) Chromosomes are localized so that smaller chromosomes are closer to the center, whereas bigger ones are localized in the outer regions.



(a)



(b)

Fig. 4. Direct validation of lamina function and compartmentalization interactions. (a) 2D spartial distribution of compartmentalization. The B compartment, which is attracted to the lamina, is concentrated near the nuclear periphery, whereas the A compartment is concentrated in the intermediate region between the inner and outer laminas. (b) The radial distribution of subcompartments demonstrates that the B compartment is attracted to the inner lamina as well and exhibits a bimodal distribution.

and attraction of B compartment pushes A compartment segments closer to the lamina boundaries as well. Nevertheless, we can observe higher enrichment of compartment B near boundaries by applying the lamina force-field.

Finally, we can model attraction of smaller chromosomes [68–70] to nucleolus which is represented by the wall $r = R_1$, by writing down the following potential,

$$E_G = G s_c \left(\sin\left(\frac{r - R_1}{\ell_G} \right) - \left(\frac{r - R_1}{\ell_G} \right)^2 \right) \tag{11}$$

where s_c takes values in the interval $[0, 1]$ and it is 1 for the chromosome with smaller length and 0 for the chromosome with a larger length, and G takes small values like 100 kJ/mol. The length parameter $\ell_G = 1\ u$ is written in the equation only to resolve issues with units and it does not exist as a parameter within the script. By making G stronger, it can result in very intense attraction to nucleolus, and thus we suggest users keep the value of this parameter low.

2.6. Optimization

For the optimization of the model there are many different options of OpenMM integrators that are included in the model: Verlet and Langevin integrator with fixed and variable steps, Brownian integrator, and AMD integrator. In principle, it is enough to run an energy minimization to produce the optimized, in terms of energy, structure. If users are not user-specified satisfied with that, they implement relaxation dynamics, by letting the system evolve for some simulation steps. The software outputs the energy and temperature of the simulation and the user can diagnose if the simulation works correctly.

2.7. Nucleosome interpolation

Nucleosome scale is included as an interpolation method in MultiMM by assuming a beads-on-the-string model of nucleosomes [5,7]. Therefore, it is assumed that DNA wraps around histones 1.65 times like

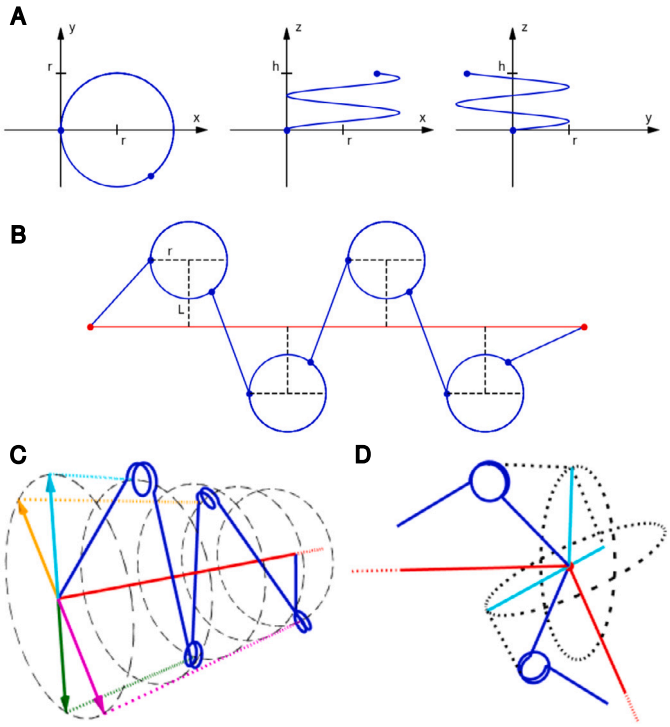


Fig. 5. Nucleosome interpolation scheme. (A) Three orthogonal projections of the generic helix which serve as a nucleosome model. Here, helix makes 1.65 turns. r and h denote the nucleosome radius and height respectively. (B) Positioning of the consecutive nucleosome along the original chromatin fragment (red segment). Here, r and L denote nucleosome radius and half of the linker length. Linker length may be constant or is variable in the "random" mode of the simulation. The chromatin interpolated with nucleosomes depicted in blue. (C) The rotation of consecutive nucleosome positioning in "zigzag" configuration. The red segment represents the chromatin fragment. The position of the fist nucleosome is determined by the first "zigzag" vector (cyan). Second vector (green) is rotated 180°. Third vector (orange) is rotated $180° + \phi_{norm}$ and the fourth (pink) another 180°. In "random" configuration the angles are randomized. (D) Junction between chromatin fragments (red segments). The last "zigzag" vector is projected onto the plane orthogonal to the next chromatin fragment and reversed. This procedure ensured relative continuity of the zigzag pattern when crossing between chromatin fragments.

a counter-clockwise helix. The user imports the ATAC-Seq signal, and the simulation places nucleosomes in a coarse-grained manner [25,71]. Therefore, regions with weak ATAC-Seq signal will correspond to nucleosome rich regions, whereas regions with high ATAC-Seq signal would be nucleosome poor. For the purpose of this study, we used the p-value of ATAC-Seq data which follows an inverse relation with the density of nucleosomes, because low a p-value means that there is a significant difference in chromatin accessibility, suggesting changes in nucleosome density or other regulatory mechanisms. Therefore, regions with high ATAC-Seq p-value correspond to nucleosome dense regions, and low ATAC-Seq signal. Consequently, we aggregate the signal so as to fit the simulation dimension by computing the average ATAC-Seq p-value for each simulation bead. The logarithm of the signal is then normalized to $[0, 1]$ and used as a basis for the number of nucleosomes to align in a given chromatin segment. The number of nucleosomes is calculated as proportional to that normalized signal with the maximum number assigned to the segments with the strongest signal and no nucleosomes are located in the segments with the weakest. The maximum number of nucleosomes allowed in a single bead is inversely proportional to the nucleosome diameter.

In order to initialize the nucleosome positions the generic nucleosome helix was first created (see Fig. 5A). This generic helix was positioned in the particular DNA fragments specified by two points in 3D space: the start p_1 and end p_2 of the nucleosome helix. The function translates the generic helix coordinates defined in standard basis of a coordinate vector space into the basis defined by a vector between the start and end point of the specified nucleosome position ($\vec{p_1 p_2}$) and two other orthogonal vectors. The second vector from this base is defined by a perpendicular component of the vector ($\vec{p_0 p_1}$) where p_0 is a point lying on an original segment corresponding to a given nucleosome. The third vector is found by cross product of the other two and all of them are normalized to form a new orthogonal basis. The coordinates of the original generic helix are then written down in the coordinates of this basis with the starting point of p_1, producing a desired helix positioning along the DNA strand (Fig. 5B). This positioning is defined by a beads-on-a-string or zigzag model with linker DNA between nucleosomes set to 3.45 times the nucleosome radius. Firstly, on each segment equidistant points are determined, one point per nucleosome. From each point the nucleosome is going to be located at half the distance of the linker DNA away from the segment by the function described above. Consequent nucleosomes are located on the opposite sides of the segment with the additional small random angle thus creating a structure roughly corresponding to the zigzag model. The angle of the last nucleosome in a segment is passed to the next assuring that the consequent nucleosomes on a different segment will not be placed on the same side of the segment (Fig. 5C and D).

The nucleosome helices are initialized based on a couple of assumptions about the nucleosome dimensions. Firstly, it is assumed in our simulation that chromatin wraps around the nucleosome 1.65 times [72]. Secondly, the ratio between the nucleosome radius and its height is assumed to be 1 based on the estimates in [73] that the nucleosome structure has a height of about 5.5 nm and a diameter of 11 nm. The ratio between the nucleosome radius and average linker DNA length is set to 1/3.45, which was based on the estimate of the average linker DNA length of 55 bp [74] and the length of a single DNA base pair of 0.34 nm. Finally, the angle ϕ_{norm}, for which the angle through which every second nucleosome is rotated in a zigzag configuration, is by default set to $\pi/5$ [75].

Furthermore, it is worth comparing the length of the nucleus radius with the scale of loops. The sizes of loops can be highly variable, from 50 kb to 200 kb [9], where each kb can be considered a 0.34 nm DNA fiber. This means that the average ratio of loop length over nucleosome radius is in the order of magnitude of $10^2 - 10^3$, because in simulation we exclude too small loops that cannot be modeled with long-range forces. In our simulation, we assume that a single loop consists of tens of beads of 0.1 u, and the radius of a nucleosome is 10^{-2} u. Considering that each loop consists of tens or hundreds of beads, the ratio remains similar to the real one.

3. Results

Validating chromatin structure remains a challenging task due to the limited availability of experimental data. Nevertheless, it is essential to demonstrate that the generated structures exhibit biophysical relevance. In this section, we analyze the chromatin structures produced by MultiMM and compare them to random models. The primary objective is to demonstrate that MultiMM, which incorporates biophysical principles, reconstructs chromatin structures more efficiently than random models. Additionally, we explore specific case studies to further evaluate MultiMM's performance in chromatin modeling.

3.1. Resulting structures

MultiMM is capable of modeling chromatin across multiple scales, ranging from nucleosomes to entire chromosomal territories. As demonstrated in Fig. 7A, at higher resolutions, MultiMM reconstructs nucleosome density through a bead-on-a-string representation. Topologically associating domains (TADs) are modeled as clusters of loops, while

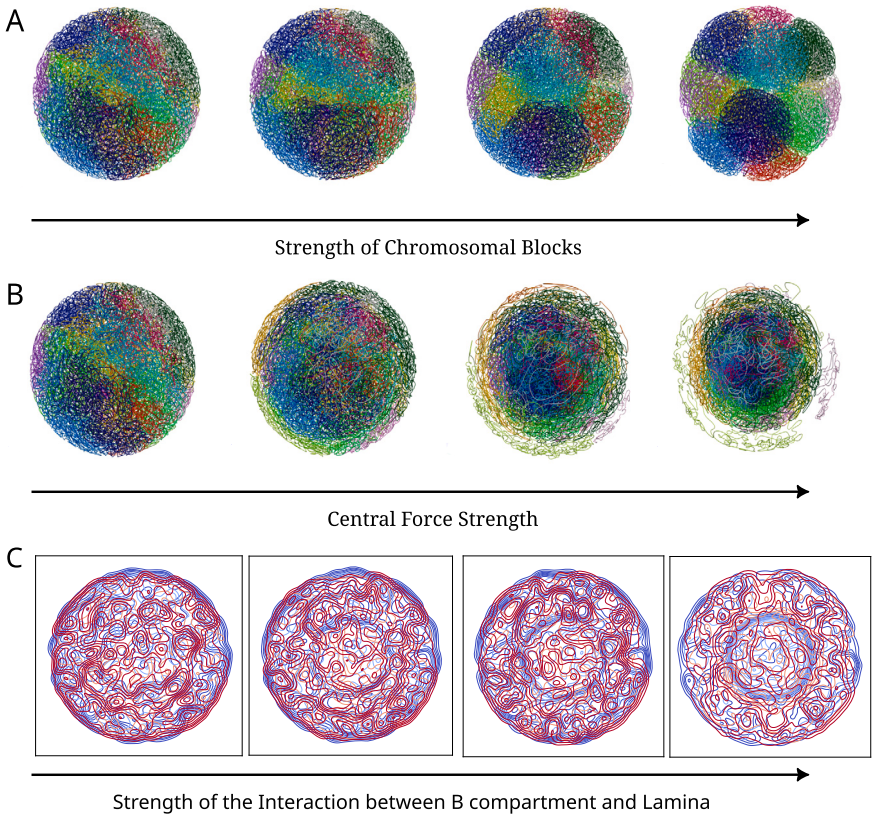


Fig. 6. Changes in chromatin structure resulting from variations in hyper-parameters for chromosomal block strength, central force strength, and B-lamina interaction strength. Each structure comprises 50,000 beads. The parameter values are: (A) chromosomal block strength: 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1} kJ/mol; (B) central force strength: 10, 100, 500, 1000 kJ/mol; and (C) B-lamina interaction strength: 100, 400, 1000, 5000 kJ/mol.

(sub)compartments are represented as groups of TADs, which are attracted to each other based on their shared (sub)compartmental block. This hierarchical modeling strategy ultimately yields the structural organization of individual chromosomes.

The output of the model provides both the structure of the entire genome and the structure of individual chromosomes. Furthermore, the model can produce a separate nucleosome-resolved structure if nucleosome interpolation is enabled by the user. The results can be visualized using the built-in functions based on PyVista [61]. Additionally, the model generates output files that provide chromosomal and compartmental color-coding.

3.2. Parameter study

Validation through biophysical concepts involves parameters that cannot be directly compared with experimental data, necessitating a qualitative approach. Such parameters include lamina-related interactions, the attraction of smaller chromosomes to the nucleolus, and chromosomal compaction. These force-field parameters are often determined using a grid search and direct observation of the spatial distribution of chromosomes and compartments within the nucleus. Direct validation against experimental data is limited by the scarcity of precise maps of compartmentalization and chromosomal territories. However, qualitative observations indicate that B compartments are attracted to the lamina (Fig. 3a), and smaller chromosomes are drawn to the nucleolus (Fig. 3b). These interactions can be tested by observing the behavior of our structure under change of the parameters of the force-field. The strength of a hypothetical force, noted as λ, should be carefully calibrated; a value that is too weak may obscure the biophysical concept, while an excessively strong potential can lead to unnatural, collapsed structures lacking biophysical relevance. Consequently, if we exclude biophysically unrealistic choices of λ we can claim that there is an interval $\lambda \in (\lambda_{\min}, \lambda_{\max})$ of acceptable choices for λ.

In Fig. 6, we illustrate the impact of varying force-field component strengths on chromatin structure. Specifically, Fig. 6A demonstrates that with a low strength of the chromosomal block, the chromosomes exhibit a loose configuration due to insufficient force-field strength. Conversely, as the strength increases, we observe a transition towards more globular chromosomal structures. Fig. 6B further highlights the effects of enhancing the central attractive force. At elevated strengths, this force draws not only smaller chromosomes but nearly all chromosomes toward the center, indicating that both excessively weak and strong forces lack biophysical relevance. Weak forces fail to significantly influence the final structure, while excessively strong forces result in collapsed and unnatural configurations. In Fig. 6C, we demonstrate that a strong attraction of the B compartment to the lamina can generate an anomalous B compartment crust adjacent to the nucleolus. Our default parameters are calibrated to achieve a balance between these two extremes for each force-field component. However, the selection of parameters may be contingent on system size and input data; thus, we encourage users to adjust these parameters as needed. Notably, our default settings have been optimized for large systems, accommodating up to half a million beads.

3.3. Validation against input datasets

The second validation approach involves a direct comparison of certain properties of the input data. For instance, when analyzing the structure of an entire chromosome, it is pertinent to assess the validity of phase separation due to compartmentalization. To this end, we simulated 20 models, each consisting of 500,000 beads, initiated from different starting configurations, and calculated the inverse distance matrix raised to the power of 3/2 for comparison with experimental

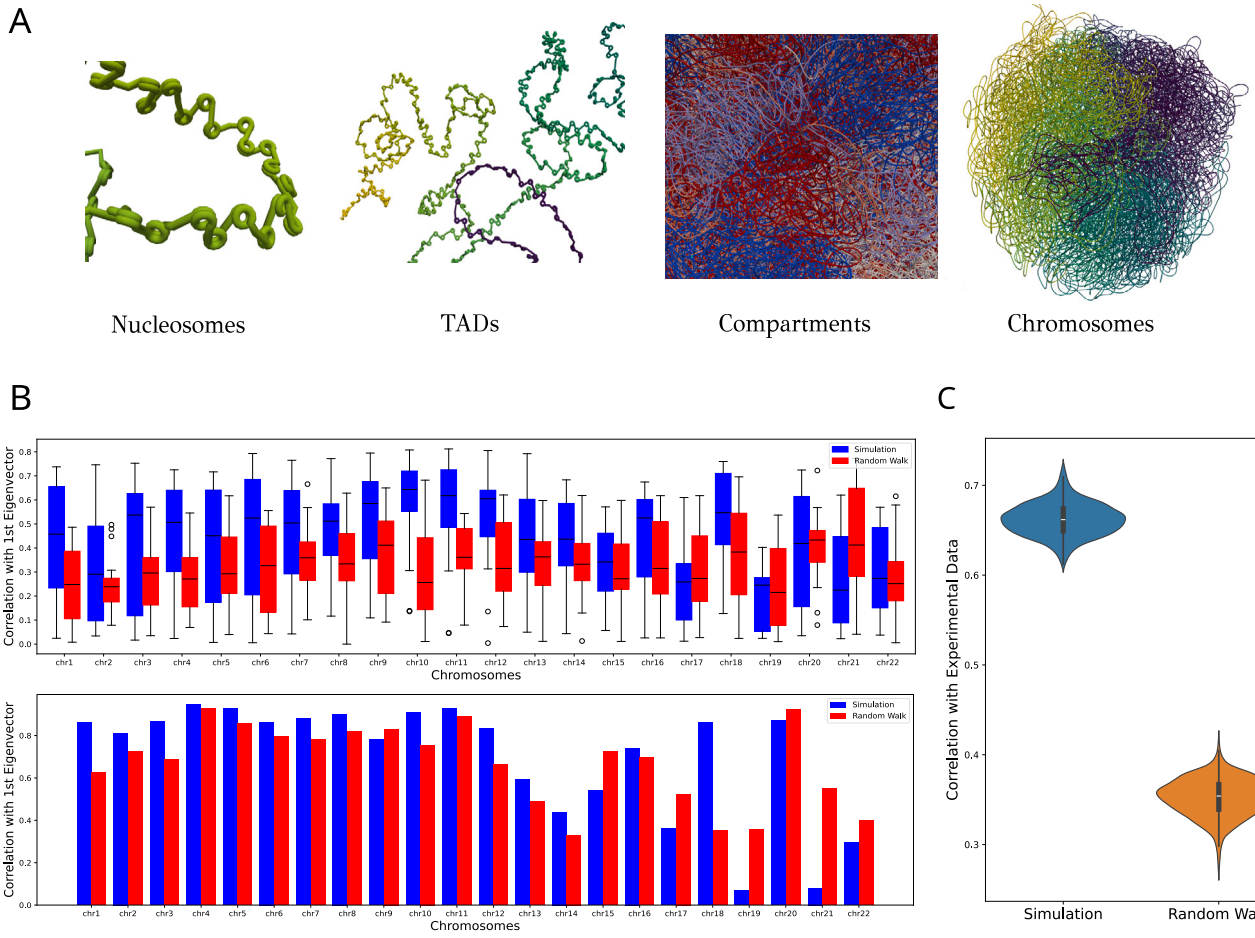


Fig. 7. Results from modeling with MultiMM. (A) Visualization of different chromatin scales using PyVista [61]. (B) The upper graph displays a plot comparing individual simulated and random inverse-distance heatmaps with corresponding experimental data. The lower graph presents a similar comparison, showcasing averages over multiple samples of heatmaps correlated with experimental data. Each structure consisted of 10,000 beads. (C) Pearson correlation coefficients between the heatmaps of structures modeled with 1,000 beads across 1,000 random 20 Mb regions of chromatin, compared to the experimental data.

datasets. We computed the first eigenvector of the correlation matrix for each heatmap, as well as the correlation for the averaged heatmap. Two experiments were then conducted: first, we evaluated the correlation between the first eigenvector of each model and the experimental data (Fig. 7B upper figure), and second, we examined the correlation between the first eigenvector of the averaged heatmap and the experimental data (Fig. 7B lower figure). Additionally, we generated 20 random walk models and repeated the procedure. The results indicate that our model exhibits statistically significant superior performance across nearly all chromosomes, particularly the larger ones. However, in smaller chromosomes, the random model demonstrated higher correlations, likely due to the presence of extensive non-interacting regions in the experimental data.

For validation at smaller scales, specific 20 Mb random regions of the chromosome were modeled. Each region likely contains dozens of TADs. To evaluate model performance at this scale, we generated 1,000 models of random regions and compared the correlations of their inverse-distance heatmaps with experimental data (Fig. 7C). By applying a stricter criterion—excluding the main diagonal—we achieved correlations around 70%, with statistically significant p-values. In contrast, random structures showed correlations below 40%.

3.4. Special cases

As described in the previous sections, MultiMM has the ability to accommodate various types of datasets. However, a critical question arises regarding how MultiMM can be configured to model specific cases effectively. To address this, it is important to examine the input data types utilized by MultiMM, which can be broadly categorized into three classes: (i) loop/interaction data, (ii) compartmentalization data, and (iii) ATAC-Seq data for nucleosome simulations. Each of these data classes corresponds to distinct force-field components. For instance, when incorporating compartmentalization data, block-copolymer forces should be activated. For modeling population-averaged 3C-type experiments, it is recommended to utilize both a list of loop interactions for short-range interactions and compartmentalization data for long-range interactions. However, compartmentalization data are not needed in case that the user prefers to model TAD scale regions avoiding the genome-wide simulation.

In specialized cases, such as Hi-C data, users may opt for hard-code long-range interactions and include them as loop interactions rather than compartments in the simulation input. In such scenarios, compartment-specific lamina interactions must be disabled, as MultiMM does not contain compartment information. Parameter adjustments can be made depending on system size. For example, if chromosomal blocks are to be included, the strength of the chromosomal blocks, k_C, should be weakened for larger structures. Other parameters that need adjustment are the boundaries of the spherical container or the range of block-copolymer interactions, but there are functions within MultiMM which tune them automatically depending on the system size (Table 3).

For single-cell data, the handling of binary interactions becomes crucial. Defining compartments from such interactions is challenging due to the inability to resolve eigenvalues and eigenvectors from binary heatmaps. Consequently, only loop interactions with fixed distances are

Table 3
Table of simulation parameters that do not take fixed values and are functions of other parameters.

Parameter Name	Symbol	Function
Small Container Radius	R_1	$R_1 = (L/(5 \times 10^4))^{1/3}$
Big Container Radius	R_2	$R_2 = 3.5R_1$
Std of Block-Copolymer Interactions	r_0	$r_0 = (R_2 - R_1)/20$
Equilibrium Loop Length	$d_{i,j}$	$d_{ij} \sim 1/S_{i,j}^{2/3}$

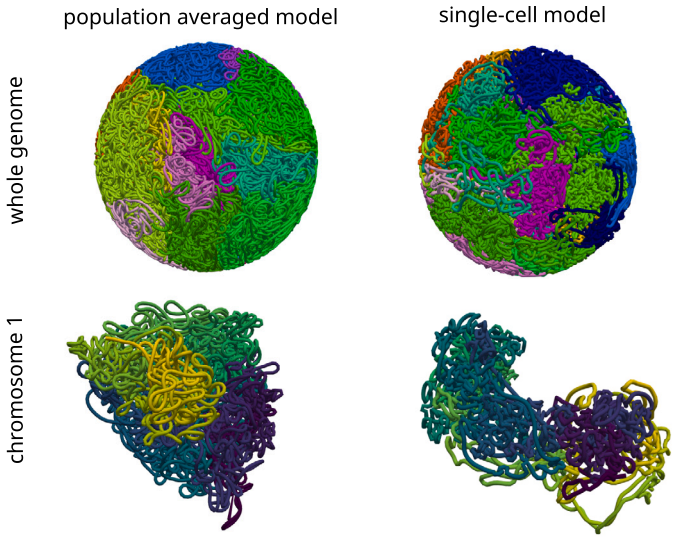


Fig. 8. Comparison between modeling with population averaged and single cell data as input. The granularity of MultiMM whole genome structure in these simulations is 100000 beads with the default parameters as input.

used, where the main opposing force is excluded volume. In this case, users may either assume that loop interactions dominate over excluded volume or seek a balance between the two, potentially leading to less compact structures. The key structural difference between single-cell and population-averaged data lies in the higher level of compaction observed in single-cell data (Fig. 8), as noise interactions, which are difficult to filter, are more prevalent in single-cell experiments.

4. Discussion

In conclusion, MultiMM offers a user-friendly platform for 3D genome-wide chromatin modeling, integrating biophysical knowledge into a unified force field applicable across various scales. This tool enables efficient simulations of chromatin architecture, allowing users to explore diverse hypotheses related to chromatin organization and force-field parameters. Additionally, the software provides intuitive visualization functionalities, and its computational efficiency is optimized through energy minimization using either GPU or multi-core CPU processing.

In its current form, MultiMM demonstrates the ability to model nucleosome intensities based on ATAC-Seq data [21,22], utilizing interpolation techniques that integrate experimental observations. On the scale of topologically associating domains (TADs), the model accepts `.bedpe` files containing loop interaction data, which are incorporated into the force field as spring or harmonic bonds, following the framework established by the "Spring model" [37]. Compartment forces are modeled using a block-copolymer approach [38], and interactions with the nuclear lamina are included to enhance the realism of the model at larger scales [40].

The primary advantage of MultiMM is its accessibility. Built on the OpenMM engine and utilizing Python, it is straightforward to install via pip and can be executed with a single command in a Linux terminal. Its flexible configuration file further simplifies customization, making the tool accessible to a broad user base. Additionally, MultiMM introduces significant improvements over the Spring model [37], including the ability to simulate much larger structures while incorporating scale-dependent potentials.

Looking ahead, we believe MultiMM represents a key step forward in the field, serving as a comprehensive tool that will catalyze further progress in chromatin research. Through its open-source framework, we anticipate that it will foster a collaborative scientific community, driving innovation and extending the boundaries of chromatin modeling. By continuing to engage with users and integrating feedback, we aim to evolve MultiMM into an even more versatile and powerful platform in future iterations.

CRediT authorship contribution statement

Sevastianos Korsak: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Krzysztof Banecki:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. **Dariusz Plewczynski:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

Authors do not acknowledge any conflicts of interest.

Data availability

For the example structures in the figures of the text, we used data from Rao et al. [13] which are publicly available in GEO (GSE63525). For the calculation of subcompartments we used the software Calder [30] (https://github.com/CSOgroup/CALDER2). Finally, we used ATAC-Seq data from ENCODE (ENCSR637XSC).

Acknowledgements

Research was funded by Warsaw University of Technology within the Excellence Initiative: Research University (IDUB) programme. This work has been co-supported by Polish National Science Centre (2020/37/B/NZ2/03757). Computations were performed thanks to the Laboratory of Bioinformatics and Computational Genomics, Faculty of Mathematics and Information Science, Warsaw University of Technology using Artificial Intelligence HPC platform financed by Polish Ministry of Science and Higher Education (decision no. 7054/IA/SP/2020 of 2020-08-28). The work was co-supported by National Institute of Health USA 4DNucleome grant 1U54DK107967-01 "Nucleome Positioning System for Spatiotemporal Genome Organization and Regulation". Special thanks to our students Michał Chęć, Cezary Rybak, Jakub Seliga and Michaił Legczylin, who tested and reviewed our code in different operating systems, and our laboratory member Michał Kadlof who helped with his insightful feedback.

References

[1] Zhou R, Gao YQ. Polymer models for the mechanisms of chromatin 3d folding: review and perspective. Phys Chem Chem Phys 2020;22:20189–201. https://doi.org/10.1039/D0CP01877E.
[2] Sikorska N, Sexton T. Defining functionally relevant spatial chromatin domains: it is a tad complicated. J Mol Biol 12.2019;432(12). https://doi.org/10.1016/j.jmb.2019.12.006.
[3] Ea V, Baudement M-O, Lesne A, Forné T. Contribution of topological domains and loop formation to 3d chromatin organization. Genes 2015;6(3):734–50.
[4] Bianco S, Chiariello AM, Annunziatella C, Esposito A, Nicodemi M. Predicting chromatin architecture from models of polymer physics. Chromosom Res 2017;25:25–34.
[5] Biswas M, Langowski J, Bishop TC. Atomistic simulations of nucleosomes. Wiley Interdiscip Rev Comput Mol Sci 2013;3(4):378–92.

[6] Zhang C, Huang J. Interactions between nucleosomes: from atomistic simulation to polymer model. Front Mol Biosci 2021;8. https://doi.org/10.3389/fmolb.2021.624679. https://www.frontiersin.org/articles/10.3389/fmolb.2021.624679.
[7] Baldi S, Korber P, Becker PB. Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. Nat Struct Mol Biol 2020;27(2):109–18.
[8] Scipioni A, Turchetti G, Morosetti S, De Santis P. Geometrical, conformational and topological restraints in regular nucleosome compaction in chromatin. Biophys Chem 2010;148(1–3):56–67.
[9] Chowdhury HM, Boult T, Oluwadare O. Comparative study on chromatin loop callers using hi-c data reveals their effectiveness. BMC Bioinform 2024;25(1):123.
[10] Hansen AS. Ctcf as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. Nucleus 2020;11(1):132–48. https://doi.org/10.1080/19491034.2020.1782024. pMID: 32631111.
[11] Goloborodko A, Imakaev MV, Marko JF, Mirny L. Compaction and segregation of sister chromatids via active loop extrusion. eLife 2016;5:e14864.
[12] Agarwal A, Korsak S, Choudhury A, Plewczynski D. The dynamic role of cohesin in maintaining human genome architecture. BioEssays 2023;45(10):2200240.
[13] Rao SS, 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(7):1665–80.
[14] Harris HL, Gu H, Olshansky M, Wang A, Farabella I, Eliaz Y, et al. Chromatin alternates between a and b compartments at kilobase scale for subgenic organization. Nat Commun 2023;14(1):3303.
[15] Hildebrand EM, Dekker J. Mechanisms and functions of chromosome compartmentalization. Trends Biochem Sci 2020;45(5):385–96.
[16] Wang L, Gao Y, Zheng X, Liu C, Dong S, Li R, et al. Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. Mol Cell 2019;76(4):646–59.
[17] Bhat P, Honson D, Guttman M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. Nat Rev Mol Cell Biol 2021;22(10):653–70.
[18] Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol 2010;2(3):a003889.
[19] Amendola M, van Steensel B. Mechanisms and dynamics of nuclear lamina–genome interactions. Curr Opin Cell Biol 2014;28:61–8.
[20] Briand N, Collas P. Lamina-associated domains: peripheral matters and internal affairs. Genome Biol 2020;21(1):1–25.
[21] Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to atac-seq data analysis. Genome Biol 2020;21:1–16.
[22] Sun Y, Miao N, Sun T. Detect accessible chromatin using atac-sequencing, from principle to applications. Hereditas 2019;156(1):1–9.
[23] Chereji RV, Bryson TD, Henikoff S. Quantitative mnase-seq accurately maps nucleosome occupancy levels. Genome Biol 2019;20:1–18.
[24] Chereji RV, Clark DJ. Major determinants of nucleosome positioning. Biophys J 2018;114(10):2279–89.
[25] Polishko A, Bunnik EM, Le Roch KG, Lonardi S. Puffin-a parameter-free method to build nucleosome maps from paired-end reads. BMC bioinformatics, vol. 15. Central: BioMed; 2014. p. 1–10.
[26] Schöpflin R, Teif VB, Müller O, Weinberg C, Rippe K, Wedemann G. Modeling nucleosome position distributions from experimental nucleosome positioning maps. Bioinformatics 2013;29(19):2380–6.
[27] Teif VB. Nucleosome positioning: resources and tools online. Brief Bioinform 2016;17(5):745–57.
[28] Li G, Cai L, Chang H, Hong P, Zhou Q, Kulakova EV, et al. Chromatin interaction analysis with paired-end tag (chia-pet) sequencing technology and application. BMC Genomics 2014;15(12):1–10.
[29] 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(11):919–22.
[30] Liu Y, Nanni L, Sungalee S, Zufferey M, Tavernari D, Mina M, et al. Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. Nat Commun 2021;12(1):2439.
[31] Ohno M, Ando T, Priest DG, Taniguchi Y. Hi-co: 3d genome structure analysis with nucleosome resolution. Nat Protoc 2021;16(7):3439–69.
[32] Bishop TC. Geometry of the nucleosomal dna superhelix. Biophys J 2008;95(3):1007–17.
[33] Korsak S, Plewczynski D. Loopsage: an energy-based Monte Carlo approach for the loop extrusion modelling of chromatin. Methods 2024.
[34] Rossini R, Kumar V, Mathelier A, Rognes T, Paulsen J. Modle: high-performance stochastic modeling of dna loop extrusion interactions. Genome Biol 11.2022;23. https://doi.org/10.1186/s13059-022-02815-7.
[35] Banigan EJ, Mirny LA. Loop extrusion: theory meets single-molecule experiments. Curr Opin Cell Biol 2020;64:124–38. https://doi.org/10.1016/j.ceb.2020.04.011. https://www.sciencedirect.com/science/article/pii/S0955067420300570.
[36] Banigan EJ, Mirny LA. The interplay between asymmetric and symmetric dna loop extrusion. eLife 2020;9:e63528. https://doi.org/10.7554/eLife.63528.
[37] Kadlof M, Rozycka J, Plewczynski D. Spring model–chromatin modeling tool based on openmm. Methods 2020;181:62–9.
[38] Jost D, Carrivain P, Cavalli G, Vaillant C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. Nucleic Acids Res 2014;42(15):9553–61.
[39] Paulsen J, Liyakat Ali TM, Collas P. Computational 3d genome modeling using chrom3d. Nat Protoc 2018;13(5):1137–52.
[40] Câmara AS, Mascher M. Consistencies and contradictions in different polymer models of chromatin architecture. Comput Struct Biotechnol J 2023;21:1084–91.
[41] Nichols MH, Corces VG. Principles of 3d compartmentalization of the human genome. Cell Rep 2021;35(13).
[42] Le Treut G, Képès F, Orland H. A polymer model for the quantitative reconstruction of chromosome architecture from hic and gam data. Biophys J 2018;115(12):2286–94.
[43] Lesne A, Riposo J, Roger P, Cournac A, Mozziconacci J. 3d genome reconstruction from chromosomal contacts. Nat Methods 2014;11(11):1141–3.
[44] Bau D, Marti-Renom MA. Genome structure determination via 3c-based data integration by the integrative modeling platform. Methods 2012;58(3):300–6.
[45] Shinkai S, Nakagawa M, Sugawara T, Togashi Y, Ochiai H, Nakato R, et al. Phi-c: deciphering hi-c data into polymer dynamics. NAR Genomics Bioinform 2020;2(2):lqaa020.
[46] Shi G, Thirumalai D. A maximum-entropy model to predict 3d structural ensembles of chromatin from pairwise distances with applications to interphase chromosomes and structural variants. Nat Commun 2023;14(1):1150.
[47] Crippa M, Zhan Y, Tiana G. Effective model of loop extrusion predicts chromosomal domains. Phys Rev E 2020;102:032414. https://doi.org/10.1103/PhysRevE.102.032414. https://link.aps.org/doi/10.1103/PhysRevE.102.032414.
[48] Maji A, Padinhateeri R, Mitra MK. The accidental ally: nucleosome barriers can accelerate cohesin-mediated loop formation in chromatin. Biophys J 2020;119(11):2316–25. https://doi.org/10.1016/j.bpj.2020.10.014. https://europepmc.org/articles/PMC7732762.
[49] Shinkai S, Onami S, Nakato R. Toward understanding the dynamic state of 3d genome. Comput Struct Biotechnol J 2020;18:2259–69.
[50] Szałaj P, Tang Z, Michalski P, Pietal MJ, Luo OJ, Sadowski M, et al. An integrated 3-dimensional genome modeling engine for data-driven simulation of spatial genome organization. Genome Res 2016;26(12):1697–709.
[51] Wlasnowolski M, Grabowski P, Roszczyk D, Kaczmarski K, Plewczynski D. cudammc: Gpu-enhanced multiscale Monte Carlo chromatin 3d modelling. Bioinformatics 2023;39(10):btad588.
[52] Paulsen J, Liyakat Ali TM, Collas P. Computational 3d genome modeling using chrom3d. Nat Protoc 2018;13(5):1137–52.
[53] Chiariello AM, Annunziatella C, Bianco S, Esposito A, Nicodemi M. Polymer physics of chromosome large-scale 3d organisation. Sci Rep 2016;6(1):29775.
[54] Sagan H. A three-dimensional Hilbert curve. Int J Math Educ Sci Technol 1993;24(4):541–5.
[55] Bountis T, Stavroulakis P. Fundamental concepts of the theory of chaos and fractals. Boca Raton, FL: CRC Press; 2006.
[56] Bountis T, Vallianatos F, Provata A, Kugiumtzis D, Kominis Y. Chaos, fractals and complexity. Springer; 2023.
[57] Bountis T. Fundamental concepts of classical chaos. Part ii: fractals and chaotic dynamics. Open Syst Inf Dyn 1997;4(3):281–322.
[58] Pascual-Reguant L, Blanco E, Galan S, Le Dily F, Cuartero Y, Serra-Bardenys G, et al. Lamin b1 mapping reveals the existence of dynamic and functional euchromatin lamin b1 domains. Nat Commun 2018;9(1):3420.
[59] Forsberg F, Brunet A, Ali TML, Collas P. Interplay of lamin a and lamin b lads on the radial positioning of chromatin. Nucleus 2019;10(1):7–20.
[60] Eastman P, Pande V. Openmm: a hardware-independent framework for molecular simulations. Comput Sci Eng 2010;12(4):34–9. https://doi.org/10.1109/MCSE.2010.27.
[61] Sullivan C, Kaszynski A. Pyvista: 3d plotting and mesh analysis through a streamlined interface for the visualization toolkit (vtk). J Open Sour Softw 2019;4(37):1450.
[62] Haddad N, Jost D, Vaillant C. Perspectives: using polymer modeling to understand the formation and function of nuclear compartments. Chromosom Res 2017;25:35–50.
[63] Jevtić P, Edens LJ, Vuković LD, Levy DL. Sizing and shaping the nucleus: mechanisms and significance. Curr Opin Cell Biol 2014;28:16–27.
[64] Pederson T. The nucleolus. Cold Spring Harb Perspect Biol 2011;3(3):a000638.
[65] Briand N, Collas P. Lamina-associated domains: peripheral matters and internal affairs. Genome Biol 2020;21(1):1–25.
[66] Attar AG, Paturej J, Banigan EJ, Erbas A. Chromatin phase separation and nuclear shape fluctuations are correlated in a polymer model of the nucleus. bioRxiv 2023 2023–12.
[67] Brunet A, Destainville N, Collas P. Physical constraints in polymer modeling of chromatin associations with the nuclear periphery at kilobase scale. Nucleus 2021;12(1):6–20.
[68] Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol 2010;2(3):a003889.
[69] Fritz AJ, Sehgal N, Pliss A, Xu J, Berezney R. Chromosome territories and the global regulation of the genome. Genes Chromosomes Cancer 2019;58(7):407–26.
[70] Meaburn KJ, Misteli T. Chromosome territories. Nature 2007;445(7126):379–81.
[71] Grandi FC, Modi H, Kampman L, Corces MR. Chromatin accessibility profiling by atac-seq. Nat Protoc 2022;17(6):1518–52.
[72] McGinty RK, Tan S. Nucleosome structure and function. Chem Rev 2015;115:2255–73.
[73] Cutter A, Hayes JJ. A brief review of nucleosome structure. FEBS Lett 2015;589:1873–3468.

[74] Bascom GD, Kim T, Schlick T. Kilobase pair chromatin fiber contacts promoted by living-system- like dna linker length distributions and nucleosome depletion. J Phys Chem 2017;8:3882–94. https://doi.org/10.1021/acs.jpcb.7b00998.

[75] Schlick T, Perišić O. Mesoscale simulations of two nucleosome-repeat length oligonucleosomes. Phys Chem Chem Phys 2009;11(45):10729–37.