



Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Mini-Review

Advancements and future directions in single-cell Hi-C based 3D chromatin modeling

Krzysztof Banecki^{a,b}, Sevastianos Korsak^{a,b}, Dariusz Plewczynski^{a,b,*}

^a Laboratory of Bioinformatics and Computational Genomics, Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

^b Laboratory of Functional and Structural Genomics, Centre of New Technologies, University of Warsaw, Warsaw, Poland



A B S T R A C T

Single-cell Hi-C data provides valuable insights into the three-dimensional organization of chromatin within individual cells, yet modeling this data poses significant challenges due to its inherent sparsity and variability. This review comprehensively explores the predominant approaches to reconstructing 3D chromatin structures from single-cell Hi-C data, positioning these methods within the broader contexts of single-cell Hi-C research and bulk Hi-C data modeling.

We categorize the modeling strategies based on their objective functions, which are framed in terms of force fields, potentials, cost functions, or likelihood probabilities. Despite their diverse methodologies, these approaches exhibit deep underlying similarities. We further dissect the basic components of these models, such as attractive restraint forces and repulsive forces, and discuss additional terms like fluid viscosity and variation penalties.

The review also critically evaluates the current state of model validation, highlighting the inconsistencies across various studies and emphasizing the need for a comprehensive validation framework. We detail common validation techniques, including the comparison of distance matrices and the assessment of contact violations.

We argue that the future of single-cell Hi-C modeling lies in integrating multiple data modalities and incorporating cell cycle trajectory information. Such integration could significantly advance our understanding of chromatin conformation dynamics during cell cycle progression and cell differentiation. We also foresee the continued growth of optimization-based and molecular dynamics approaches, supported by general molecular dynamics toolkits.

1. Introduction

Since the introduction of the Hi-C data in 2009 [1], there has been a significant opportunity for accurate three-dimensional (3D) genome structure reconstruction. A multitude of early methods for 3D structure reconstruction soon followed [2–11]. However, challenges arose in reconstructing the 3D genome structure from bulk Hi-C, typically averaged over millions of cells. This averaging process inherently fails to capture the heterogeneity of underlying chromatin structures and some of their topological properties. Early methods often addressed this by attempting to deconvolute the Hi-C contact matrix or using Monte Carlo sampling to simulate chromatin heterogeneity. Nonetheless, the inherent inability of bulk Hi-C data to resolve some of the ambiguous conformations has been noted for knotting patterns [12], chirality [13,14] or ergodicity of interaction frequencies [15]. Moreover, unphased Hi-C data is challenged by the ambiguity of contact assignment between homologous chromosomes in diploid organisms [16]. The introduction of the single cell Hi-C (scHi-C) data in 2013 [17], while still facing several of the same issues inherent to bulk Hi-C analysis, opened new possibilities for the 3D chromatin conformation reconstruction methods. scHi-C

data can provide valuable insights into the diverse structural landscape at the single-cell level. This new data modality enables researchers to relate probabilistic chromosome conformations with the averaged view from millions of cells [17]. Among other advances, it allows testing hypotheses regarding whether certain chromatin structural features, such as Topologically Associated Domains (TADs), are present in individual cells or are emergent properties resulting from the averaging process [18].

Analyzing 3D genomes at single-cell resolution offers a unique opportunity to explore the heterogeneity and dynamics of higher-order chromatin structure and function. Nevertheless, the inherent sparsity, noisiness, and high dimensionality of scHi-C datasets present significant analytical challenges. Specialized computational methods have been developed to address these challenges, encompassing data processing, dimensionality reduction, contact map interpolation, 3D genome structure modeling, and other downstream analyses, such as identifying 3D genome features at various scales. These algorithms have been described in several review papers [19,20], which analyze workflows for scHi-C data processing, pretreatment methods, missing data imputation, scHi-C embeddings, clustering methods, and pseudo-time series analysis.

* Corresponding author.

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

<https://doi.org/10.1016/j.csbj.2024.09.026>

Received 6 August 2024; Received in revised form 27 September 2024; Accepted 29 September 2024

Available online 3 October 2024

2001-0370/© 2024 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

To date, very few deep learning algorithms have been used in 3D chromatin structure reconstruction. This scarcity is likely due to the lack of high-quality chromatin structures that can serve as “ground-truth” for training deep learning models, rendering supervised learning approaches particularly challenging. However, deep learning algorithms have been successfully applied to predicting Hi-C contact maps from genomic [21] and epigenomic [22] sequences and for single-cell contact imputation [23]. Despite these successes, applying deep learning methodologies to 3D chromatin structure reconstruction remains in its infancy and requires further development to become applicable to single-cell data [24].

In this review, we focus on structure inference methods based on single cell Hi-C data along with their differences and similarities with methods developed for bulk Hi-C. To our knowledge, computational approaches focusing on multi-scale structure modeling in single-cell Hi-C have not been thoroughly reviewed yet. Here, we review the methods published thus far for 3D chromatin reconstruction from scHi-C data (see Table 1). Several approaches do not fall distinctly into either the bulk Hi-C or scHi-C categories. We conducted a comprehensive search for all available methods designed to analyze single-cell or single-nucleus Hi-C data. To facilitate a comparison of validation strategies, we included only those methods that provided a validation or testing section demonstrating performance on an available scHi-C dataset. Consequently, certain closely related methods were excluded, such as ShRec3D [25] or ShRec3D+ [26]. Ultimately, we identified and included 12 methods published to date in this review.

Table 1 provides a summary of the methods reviewed, including information on the general techniques employed for chromatin structure reconstruction, the programming languages used for implementation, the single-cell datasets analyzed, and links to the respective source codes. The specific techniques underlying each method are elaborated upon in the subsequent sections. C++, Python and MATLAB are the most commonly used programming languages for developing these methods. The scHi-C datasets frequently employed in single-cell 3D chromatin reconstruction include those from Nagano et al. 2013 [17], Nagano et al. 2017 [27], Stevens et al. 2017 [28], Bonev et al. 2017 [29], Tan et al. 2018 [30], Tan et al. 2019 [31], Gassler et al. 2017 [18] or Flyammer et al. 2017 [32]. Other datasets of relevance, particularly in related areas such as scHi-C clustering and contact imputation, include those from Ramani et al. 2017 [33], Lee et al. 2019 [34], and Collombet et al. 2020 [35], among others. For more information about single cell Hi-C datasets used in those fields, we direct the reader to reviews such as [36]. The majority of the methods are publicly accessible via the provided source code links.

In conclusion, this review highlights the current advancements and challenges in 3D chromatin structure reconstruction using scHi-C data, providing a comparative analysis of available computational tools with particular emphasis on their underlying techniques and validation strategies. By presenting this overview, we aim to facilitate future research and development in the rapidly evolving field of single-cell genome architecture.

2. Methods overview

Constructing models from bulk Hi-C data presents several limitations, mainly due to the fact that it typically reflects chromatin contacts averaged across millions of cells. Since the chromatin conformational features of a population cannot be easily equated with those of individual cells [45], bulk Hi-C modeling strategies often struggle to capture the intrinsic heterogeneity of chromatin conformation dynamics. To overcome these challenges and in the hope of moving toward the determination of chromatin structures in individual cells [17], a distinct set of methods has been developed to model chromatin conformations using scHi-C data. These methods are specifically designed to address the primary challenge of scHi-C data—its inherent sparsity—by employing techniques that effectively mitigate this limitation.

Several such methods have already been implemented (see Table 1), incorporating various approaches to the problem. Most of these methods use some form of the scoring function, which is then optimized by Simulated Annealing protocols or Gradient Descent optimization (e.g. [40,42]). Others define the posterior Bayesian probability function and apply Hamiltonian Monte Carlo algorithms to draw models from the distribution [38]. Still, others opt for generic Molecular Dynamics simulations [28]. Most of those methods are written in commonly utilized programming languages and are relatively accessible and usable tools for studying single-cell conformations at a single-cell resolution.

2.1. Bulk Hi-C vs scHi-C models

In general, 3D chromatin models designed for handling single-cell Hi-C data use methodologies similar to those developed for bulk Hi-C. We refer to models that use bulk Hi-C data as their primary input as “bulk Hi-C models” and those that primarily rely on single-cell data as “scHi-C models”. Some methods claim to be capable of working with either binary single-cell Hi-C matrices or quantitative bulk matrices that preserve contact frequencies [25]. Others use both modalities simultaneously [40], blurring the boundary between bulk Hi-C and scHi-C methods. Nevertheless, some general differences exist between models that use these two kinds of data.

Compared to methods designed for bulk Hi-C data, a distinctive feature of scHi-C modeling is that the exact relationship between averaged contact density and averaged 3D Euclidean distance is less pronounced. In contrast, in chromatin modeling based on bulk Hi-C data, the precise power law relationship between these parameters is widely discussed [46,47]. This relationship is often assumed to take the form $d_{ij} = F_{ij}^{-\alpha}$, with d_{ij} and F_{ij} being the Euclidean distance and contact frequency, respectively, between a given pair of loci i and j . Parameter $\alpha > 0$ is usually either taken as a constant parameter in simulations [2,48,49] or estimated/optimized in some way [4,25,26,50,51]. The discussion of this parameter, first postulated by [1], can vary depending on the studied organism or even on the genomic scale and resolution [50,52,53] and is often entirely omitted in the field of scHi-C methods. This omission is likely due to the extreme sparsity of scHi-C data, which makes estimating this parameter particularly challenging. At the same time, the limited number of contacts in a single cell enables modeling strategies that, in principle, can accurately account for each individual contact rather than relying on contact frequencies—strategies that might be computationally infeasible given the volume of contacts in bulk Hi-C datasets. Typically, this is achieved by applying a specific force-field constraint to each contact in scHi-C data. These approaches are discussed in detail in the following section. Similarly, while Hi-C data is often represented as a matrix with values representing any number of contacts between a given pair of loci, scHi-C matrices are often binary. Some models apply binarization to the extent of accepting the loss of information inherent in situations where more than one contact is present between a pair of loci at a given resolution [39].

When comparing bulk Hi-C and scHi-C modeling strategies, there is a subtle yet notable distinction in the general approaches to 3D chromatin structure reconstruction. In a previous review, reconstruction methods were classified into two categories: data-driven models and de novo ensembles [54]. Data-driven models focus on relying solely on experimental contact data without making assumptions about the genomic mechanisms that shape chromatin structure. In contrast, de novo approaches test specific biological hypotheses, such as whether loop extrusion (LE) mechanisms can generate chromatin structures that agree with Hi-C maps. While both subgroups are well represented in bulk Hi-C modeling [54], scHi-C approaches have, so far, been predominantly data-driven. All the methods reviewed here can be broadly classified as data-driven models.

Despite the challenges, de novo modeling of biological processes with scHi-C data remains an intriguing possibility. Currently, it remains unclear whether stochastic loop extrusion models [55–58] can be suc-

Table 1

Overview of computational tools for predicting 3D chromatin structure from scHi-C contact data. For each method, we provide details on the primary reconstruction technique, the implementation programming language, the single-cell datasets utilized in the original study, and links to the corresponding source code.

Method name	Technique	Language	Single cell dataset	Source code link
Nagano et al. 2013 [17]	Simulated Annealing	—	Nagano 2013	—
MBO [37]	Gradient Descent	Matlab	Nagano 2013	http://folk.uio.no/jonaspau/mbo/
ISDHi-C [38]	Hamiltonian Monte Carlo	Python	Nagano 2013	https://github.com/michaelhabeck/isdhic
RPR [39]	Recurrence Plot-based Reconstruction	Matlab	Nagano 2013	—
NucDynamics [28]	Simulated Annealing	Python	Stevens 2017	https://github.com/tjs23/nuc_dynamics
SIMBA3D [40]	Gradient Descent	Python	Stevens 2017	https://github.com/nerettitlab/SIMBA3D
SCL [41]	Metropolis-Hastings and Simulated Annealing	C++	Nagano 2013, Stevens 2017, Tan 2018	http://dna.cs.miami.edu/SCL/
Wetterman et al. 2020 [13]	Molecular Dynamics	Python	Stevens 2017	—
Si-C [42]	Gradient Descent	C++	Stevens 2017	https://github.com/TheMengLab/Si-C/
LJ3D [43]	Metropolis-Hastings and Simulated Annealing	C++	Bonev 2017	http://dna.cs.miami.edu/LJ3D
DPDChrom [44]	Dissipative Particle Dynamics	Python	Flyamer 2017, Gassler 2017	https://github.com/polly-code/DPDchrom
Rothörl et al. 2023 [14]	Molecular Dynamics	Python	Tan 2018, Tan 2019	https://gitlab.rlp.net/3d-diploid-chromatin/simulation-code/

cessfully adapted for single-cell modeling. Currently, most single-cell structures are inferred using population-averaged 3C-type data. These models rely on simplified assumptions about the dynamics of loop extrusion factors (LEFs) such as cohesins and condensins, and barrier elements like CTCFs, many of which have been observed and validated in single-cell experiments [59]. Key assumptions include the random walk motion of LEFs, the barrier activity of CTCFs, and the rebinding rates of LEFs, which facilitate system mixing. Through multiple iterations of these stochastic simulations, ensembles of 3D structures can be generated, each potentially corresponding to different single-cell conformations. While single-cell data can serve as a validation tool for these structures [60,61], it is rarely used as input for these models.

Methods for 3D chromatin reconstruction can also be categorized into two other groups: consensus and ensemble methods [46], with many representatives in each category, though most likely no strict division exists [54]. Consensus methods (e.g. [25,62–65]) aim to infer a single chromatin structure representative of a given Hi-C dataset. However, this approach usually neglects the fact that a bulk Hi-C matrix typically represents an averaged view from millions of cells, thereby overlooking the substantial heterogeneity of single-cell conformations. On the other hand, ensemble methods (e.g. [66–69]) attempt to capture cell-to-cell structural variability by inferring an ensemble of structures, for example, through deconvolution of the bulk Hi-C matrix [70–72]. In the case of single-cell Hi-C data, there is no need to deconvolve averaged data into ensembles since it already represents a single nucleus. Possibly for this reason, scHi-C models are typically consensus methods, though exceptions exist [40].

In this review, we introduce a new classification of modeling strategies. We categorize scHi-C models into two primary approaches: potential optimization-based and probabilistic methods. This division is not absolute, nor does it encompass all possible modeling strategies. Consequently, some of the reviewed methods are placed in a separate category, as they exhibit characteristics that do not fit neatly into either the potential optimization-based or probabilistic approaches.

2.2. Potential optimization-based approaches

Data-driven modeling from 3C-type data is often conceptualized using polymers and beads-on-a-string models to represent chromatin. The number of beads or monomers in such a polymer is determined by the simulation resolution and the size of the genomic region being modeled. The modeling strategy typically begins with an initial conformation, usually random, which is then sequentially refined throughout the simulation to satisfy a given set of conditions. The typical approach to reconstructing chromatin conformation from scHi-C or bulk Hi-C data involves creating a general potential/energy function. Depending on the method, this function comprises several additive components, such as a

general repulsive potential, an attractive or stabilizing contact potential, and others. The repulsive potential ensures the fulfillment of the excluded volume condition and is often applied to all pairs of monomers in the polymer. Contact restraints ensure that the modeled conformation converges to a state where the loci connected by contacts are in close proximity. Additionally, some methods include extra components in their potential functions to enhance other properties of the chromatin models or to act as penalty terms, thereby preventing the chromatin from adopting certain undesired properties.

The functions used to construct potentials in the field of single-cell modeling are varied. Some are fundamental, derived from basic physical principles, while others are approximate potentials, applied to model complex forces that often lack analytical solutions to Newton's second law. One widely used fundamental potential in single-cell modeling is the harmonic bond potential, expressed as $V_{\text{res}} = k(r - r_0)^2$, which describes oscillatory behavior similar to a spring with stiffness k and equilibrium distance r_0 [73]. In atomic physics, this potential is frequently employed to maintain atoms at their equilibrium distances. Another commonly used potential in molecular dynamics simulations is the Lennard-Jones potential, expressed as $V_{\text{LJ}} = \epsilon \left[\left(\frac{\sigma_1}{r} \right)^a - \left(\frac{\sigma_2}{r} \right)^b \right]$. It approximates atomic interactions by both mimicking attractive forces and preventing atomic overlap [74,75]. While often used as the repulsive component in force fields (e.g. [38]), it can also model both repulsive and attractive restraint forces between molecules [43]. Both harmonic and Lennard-Jones potentials are considered stringent, meaning they can impose arbitrarily large penalties as their potential function values grow without bound (see Fig. 1C-D). However, in some cases, model developers may prefer less stringent alternatives, opting for potentials that impose a finite penalty [41]. The most common of these less stringent functions are the Gaussian and sigmoid potentials. Gaussian potentials [76–78], $V_G = C \exp\left(-\frac{(r-r_0)^2}{2\sigma^2}\right)$, are favored for their smoothness in simulating complex interactions, such as block-copolymer interactions in chromatin modeling [79]. Sigmoid potentials, $V_s = \frac{V_0}{1 + \exp[-(r-r_0)/\sigma]}$, are particularly useful for modeling state transitions and are effective in accounting for excluded volume effects [41,42].

In the context of scHi-C modeling, one of the simplest approaches to constructing a potential function was introduced by Wetterman et al. 2020 [13]. Their method, inspired by approaches borrowed from the field of protein folding simulations and referred to as G δ -models [80], utilized a “minimal” potential function comprising only a Gaussian-shaped excluded volume component and harmonic components for adjacent monomers and contacts. Similarly straightforward was the approach taken by Nagano et al. 2013 [17], where the potential function included a general repulsive harmonic force, a flat-bottom harmonic restraint force, and an additional repulsive harmonic force applied to

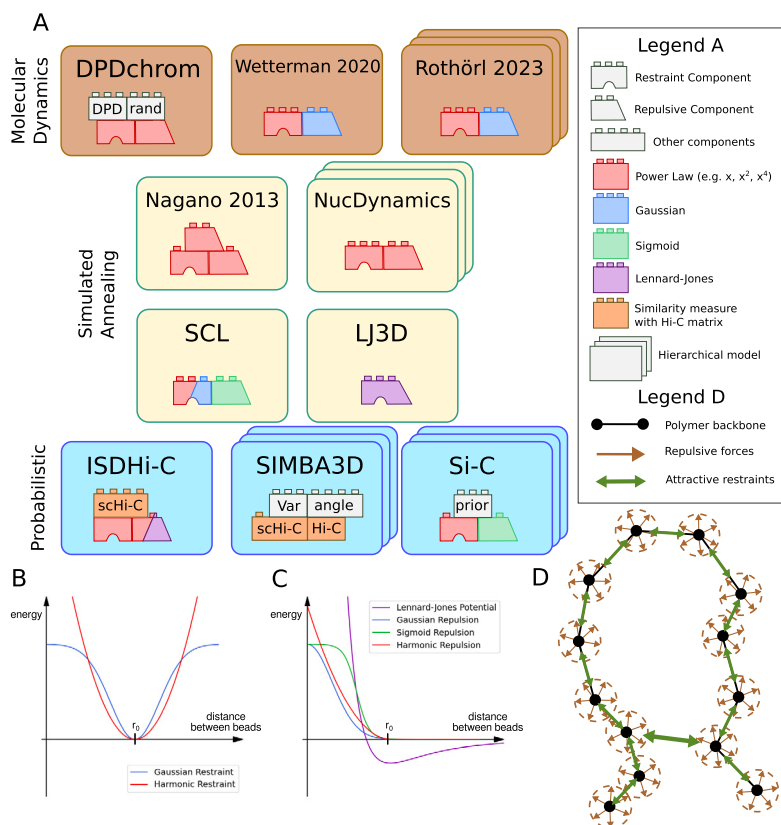


Fig. 1. (A) Schematic representations of optimization-based and probabilistic models described in this review. Technical details have been omitted, focusing instead on the fundamental principles and components of the models' general potential or probability functions. In LJ3D, the Lennard-Jones potential serves as both the restraint and the repulsive force. DPDchrom uses two additional force field components related to Dissipative Particle Dynamics: one related to viscosity and one random. SIMBA3D adds two components that penalize model variance and angles between particles. Si-C multiplies the likelihood function by the prior probability. (B) Examples of harmonic (power law) and Gaussian restraint potential functions. Target distance x_0 is indicated on the plot. Different strengths of those potentials might be used for backbone and scHi-C contact interactions. The X-axis represents the distance between particles. (C) Examples of repulsive forces responsible for the excluded volume effect in chromatin models. Repulsive potentials are often truncated and set to zero for distances exceeding a specified threshold. The X-axis represents the distance between particles. (D) Schematic representation of a beads-on-a-string polymer. Repulsive forces are shown as brown arrows. Attractive restraints, both along the polymer backbone and for one inter-bead contact, are shown as green arrows.

non-interacting regions in the ensemble Hi-C dataset, with backbone particles restrained by a strict upper distance limit.

NucDynamics [28] builds on the general modeling strategy presented by Nagano et al. (2013). This study applied a similar harmonic, flat-bottom force field to backbone monomers and contact pairs, using a general repulsive force with maximum strength 25 times weaker than the restraint force, akin to Nagano et al. The authors introduced several adjustments, including a constant optimal range for backbone particles, dependent on the number of binned contacts between loci using a power-law relationship with $\alpha = 1/3$. The method also uses multiple resolutions for hierarchical modeling, starting from the lowest resolution, and adjusts the ratio between the repulsive component and other forces during the simulation, starting close to 0 and converging to 1/25 of the restraint force strength following a sigmoid function. This adjustment helps the model avoid local minima during simulation by allowing particles to pass close to each other at the beginning of each phase. This algorithm also facilitates the modeling of single diploid human cells with Dip-C software [30].

Two other methods in this category, SCL [41] and LJ3D [43], use cubic-lattice representations of beads. In these studies, repulsion and restraint forces are framed more as cost functions rather than potentials or force fields, though the general principle remains the same. Both studies first “smooth out” sparse single-cell Hi-C matrices using a Gaussian kernel function. SCL assigns a quadratic cost function to pairs of loci with the strongest smoothed signals and a less stringent reverse Gaussian function to those with weaker but still relatively strong signals (see

Fig. 1B). The weakest signals are associated with a sigmoid penalty for close distances, interpreted as a repulsive force. LJ3D simplifies this into a Lennard-Jones potential function, scaled by the power of the smoothed Hi-C matrix signal.

Inspired by dissipative particle dynamics (DPD), Kos et al. 2021 [44] introduced a set of forces within their modeling strategy, DPDchrom. This method uses four forces: soft repulsion force, which linearly decreases up to a certain distance and is absent for more distant pairs of loci; the dissipative force of viscous friction proportional to the current particle velocity, a random force component, and an elastic force dependent on the difference between the current and optimal native-state distances.

Expanding on Wetterman et al. 2020, Rothörl et al. 2023 [14] published a method with similar potential functions but introduced additional features. Their model operates hierarchically with ambiguous diploid contact resolution at each simulation resolution, and both contact restraint and excluded volume potentials are gradually reinforced, echoing concepts from NucDynamics. Both Wetterman et al. 2020 and Rothörl et al. 2023 use GPU-friendly general Molecular Dynamics simulation software HOOMD-blue [81,82].

In general, the toolbox used in the reviewed scHi-C modeling strategies and the general approaches taken by the studies in this category share several similarities. Most methods define a cost function, potential/energy function, or force field, typically consisting of additive components (see Fig. 1A). These components generally include an excluded volume potential to prevent chromatin beads from being too close and a

restraint potential/force to ensure that particles forming the backbone of the chromatin model and monomers connected by scHi-C contacts are attracted to each other. Functions used in constructing the potential are usually harmonic, Lennard-Jones, Gaussian, or sigmoid functions in various configurations. Additional forces and potentials are often incorporated to enhance desired simulation properties, but the core features of excluded volume and restraint potentials are common across nearly all reviewed scHi-C models.

2.3. Probabilistic and Bayesian approaches

Unlike methods that rely on force-field interpretations for chromatin modeling, at least three published scHi-C methods frame the problem in probabilistic terms, specifically through Bayesian inference or maximum likelihood estimation. These methods are ISDHi-C [38], SIMBA3D [40] and Si-C [42]. Given that Bayesian and probabilistic approaches form a substantial portion of Hi-C-based methods for chromatin structure reconstruction [4,10,83], it is not surprising to see representatives of this approach in the single-cell field as well.

For example, ISDHi-C defines the problem in terms of a posterior distribution, expressed as:

$$\mathbb{P}(X, \theta | D, I) = \frac{\mathbb{P}(D | X, \theta, I) \mathbb{P}(X, \theta | I)}{\mathbb{P}(D | I)}$$

where X represents model coordinates, D denotes input data (such as scHi-C), I stands for prior information, and θ represents model parameters. In ISDHi-C, the model parameters θ are estimated alongside the model coordinates X , but they could also be assumed a priori and thus counted as part of the prior information I . Despite this probabilistic framing appearing quite different from force-field methods, the prior probability function $\mathbb{P}(X, \theta | I)$ in ISDHi-C is defined using an exponential function of a general potential function, similar to those used in optimization methods. This potential consists of two components: a harmonic restraint potential applied to consecutive backbone particles, and a “nonbonded” potential addressing the excluded volume effect, presented as either a quartic or Lennard-Jones potential function. In Monte Carlo simulations, the normalizing part $\mathbb{P}(D | I)$ of the posterior probability is often unnecessary; thus, the focus is on specifying the probability of the input data D given the model coordinates, parameters, and possibly prior information. ISDHi-C offers two alternative functions for measuring the compatibility between chromatin structure and scHi-C data: a Gaussian with a flat plateau and a logistic function, though other similarity functions used in the field could be applied as well.

Probabilistic framing often involves Monte Carlo simulations, but some methods, such as SIMBA3D [40], use direct optimization of the posterior probability distribution function. SIMBA3D constructs a posterior energy function with four main components: the negative log-likelihoods of model coordinates given single-cell Hi-C and bulk Hi-C data, ensuring that no significant violations occur between the chromatin model and both scHi-C contacts and general bulk Hi-C map patterns. The likelihood function assumes that the number of contacts follows a Poisson distribution with intensity related to a predefined power of the distance between chromatin beads. The remaining two terms are penalty terms: one penalizes variations in distances between adjacent points, and the other penalizes deviations from alignment in a straight line, thus controlling the stiffness and smoothness of the resulting conformation. The posterior energy function is optimized using a gradient-based approach, with a hierarchical multiscale optimization process that starts from the lowest resolution and interpolates model beads before initiating the next optimization problem.

Si-C [42], the final method in this category, also frames the chromatin inference problem in terms of maximum likelihood, presenting a similar approach to SIMBA3D. Here, the posterior probability is a product of probabilities $\mathbb{P}(r_{ij} | c_{ij})$ for a given distance r_{ij} and contact information c_{ij} between pairs of chromatin loci. Using Bayes’ theorem, the authors optimize the posterior probability $\mathbb{P}(c_{ij} | r_{ij}) \mathbb{P}(r_{ij})$. The prior

probability $\mathbb{P}(r_{ij})$ assumes a uniform distribution of beads, proportional to the surface area of a sphere and, therefore, to r_{ij}^2 . The probability $\mathbb{P}(c_{ij} | r_{ij})$ varies based on whether contacts exist between a given pair of monomers. For pairs without contact, this probability is modeled as a sigmoid function centered at a predefined distance r_0 , favoring larger distances. For pairs with at least one contact, the probability is proportional to the c_{ij} -th power of the difference between distance r_{ij} and optimal r_0 , and zero for greater distances. This part makes pairs of loci with the most contacts more probable to be closer to each other and absolutely unlikely when the distance is larger than r_0 . Therefore, the first term can be interpreted as an analogous substitute for the excluded volume potential used in other models. Similarly, the second term mirrors the restraint force. Optimization of this probability involves taking the negative logarithm, resulting in potential energy similar to those in other methods with two components responsible for successful chromatin modeling simulation. Similarly to the SIMBA3D method, the potential is optimized in a hierarchical manner, with interpolation taking place until the desired resolution is achieved.

Though probabilistic and Bayesian methods might initially appear quite different from optimization-based approaches, they share substantial similarities. Ultimately, all methods from both categories construct some form of a grand cost function, potential energy, or posterior distribution to describe the desired chromatin model in accordance with input single-cell contact data. These functions almost always include two main components: one addressing the excluded volume effect to prevent chromatin model particles from overlapping and the other acting as a restraint force to ensure that pairs of loci with scHi-C contacts are more likely to be close to each other in 3D space. Various methods add additional terms to these basic functions and use diverse optimization strategies, such as simulated annealing, MCMC sampling, or gradient descent. Nevertheless, the fundamental approach to building chromatin models remains fairly constant across different methods.

2.4. Other scHi-C structure reconstruction methods

Finally, it is worth mentioning two methods that do not neatly fit into either potential optimization-based or probabilistic modeling categories and were not featured in Fig. 1: Manifold Based Optimization (MBO) [37] and Recurrence Plot-based Reconstruction (RPR) [39].

The MBO approach builds on classical multidimensional scaling (MDS). Using simple eigen decomposition, it derives the optimal chromatin model reconstruction from a Euclidean distance matrix (EDM) and the Gramm matrix. Essentially, MBO solves an optimization problem based on the difference between EDM matrices and known distances from scHi-C experiments. This method does not define a general potential or posterior probability function but retrieves the optimal structure directly from contact information and mathematical concepts instead.

The other method, RPR, is also highly theoretical. It draws inspiration from the concept of recurrence plots [84] and their resemblance to the single-cell Hi-C matrices. A weighted graph of neighboring beads is created based on the input scHi-C matrix treated as a recurrence plot. From a weighted graph, a full distance matrix is obtained by the Dijkstra algorithm, and from a distance matrix, the MDS algorithm enables the final chromatin structure to be obtained. Like MBO, RPR does not require a specific force field or potential function for its simulation but relies on established theoretical algorithms from physics and mathematics.

Despite their ingenuity, these models are relatively rare in single-cell Hi-C modeling. They also seem to receive relatively less follow-up as they were one of the earliest single-cell Hi-C models developed. Those two methods happen to be also the ones with the strongest association with the MDS reconstruction approach. RPR uses MDS directly, and MBO expands on one of the MDS strategies. Subsequent methods have predominantly employed optimization-based or probabilistic approaches, forming the majority of methods developed so far. It is unclear whether the same conclusion holds for the bulk Hi-C modeling field, but

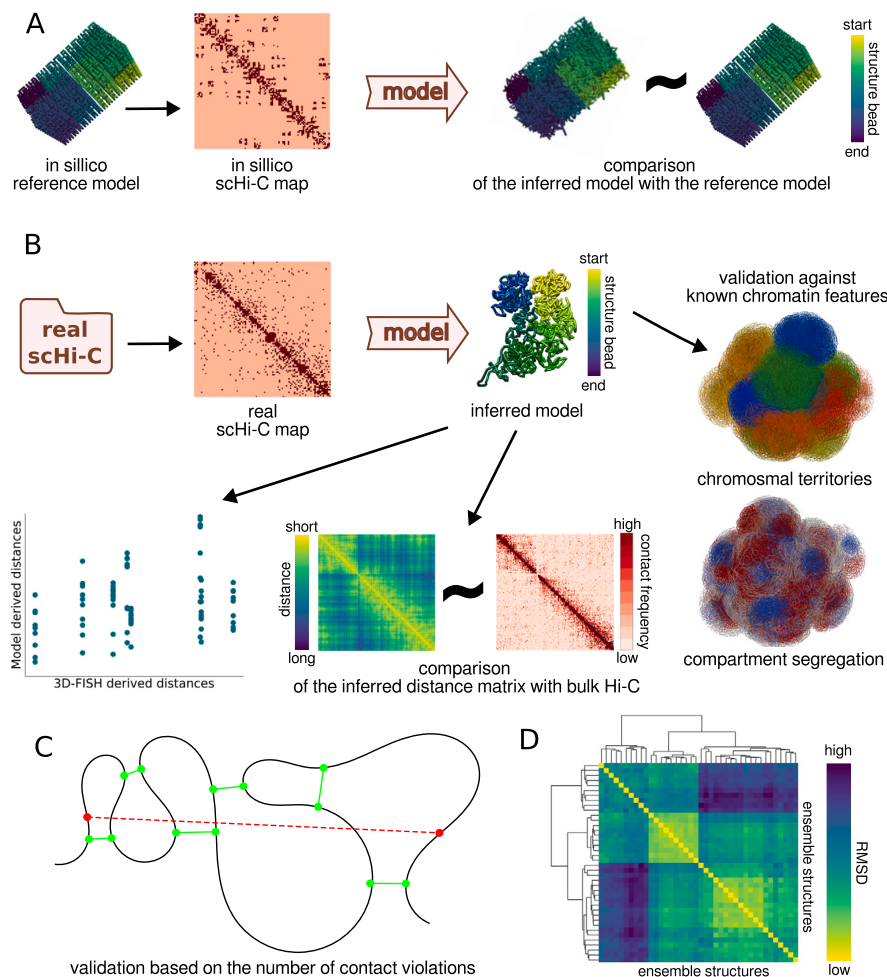


Fig. 2. Validation strategies applied in the field of single cell Hi-C modeling. (A) Validation against artificial in silico model. In this approach, an artificial single-cell Hi-C map is created from an initial structure using probabilistic assumptions about scHi-C contacts. The modeled structure is then compared with the original structure using metrics such as RMSD. (B) Validation based on real scHi-C data. Structures derived from scHi-C matrices undergo three main types of validation: calculation of the correlation between model distances and distances obtained from independent experiments, such as 3D-FISH, comparison of the structure's distance matrix with bulk Hi-C data or validation against known chromatin features, such as chromosomal territories or compartment segregation. (C) Testing the model by counting the number of violated contacts (red dashed line) against the number of successfully resolved contacts (green solid lines). (D) Clustering analysis enables validation of the model convergence. The clusters might represent local optima in the structure space or different cells if ensembles from different scHi-C maps are compared [40].

at least in the case of scHi-C modeling it makes further development of methods directly or indirectly using the MDS approach uncertain.

3. Validation strategies

Validation of the three-dimensional chromatin models is a difficult and problematic task [38,85,86]. This seems to be true regardless of the kind of genomic data used as an input to the simulation, and the models for chromatin reconstruction from scHi-C data are no exception. One of the primary reasons for this is the lack of high-resolution ground truth conformations to which the scHi-C models can be compared.

Similarly to models operating on bulk Hi-C data [46], scHi-C models are typically validated using two principal approaches. The first approach involves validation with artificial, simulated datasets [38,39], allowing controlled testing of models in a known environment (Fig. 2A). For instance, the RPR model [39] demonstrates this approach by initially testing their method with the Rössler attractor [87], a system easily simulated through numerical integration of differential equations. In this approach, the reconstruction accuracy is typically established with measures such as Root Mean Square Deviation (RMSD) or Modified Jaccard Index [44]. This approach is also helpful for assessing model behavior under varying conditions, including gauging the robustness to various noise levels or the presence of outliers. The second

approach involves validating models against independent experiments such as three-dimensional Fluorescent In Situ Hybridization (3D-FISH) [38,41,42,53,88]. FISH datasets provide detailed information on the spatial distribution of chromatin structures, making them valuable for assessing the accuracy of chromatin organization models in 3D. Possibly, the easiest way to compare simulation results with this kind of external data is to compute the Pearson correlation coefficient between bead distances obtained from the model and distances from the other experiments, an approach taken for instance by [41–43] (Fig. 2B). The most common dataset used so far in scHi-C modeling seems to be the GAM dataset [89], which directly provides a set of known distances. Experimental oligopaint data can also be used in this context [90,91].

The FISH experiments are considered reliable for the task of validation [46]. However, calculating the correlation between distances is not the only way the models can be validated against known features of the chromatin structure. For example, the method can be checked whether the resulting structures preserve chromosomal territories, Rabl configuration (when centromeres and telomeres cluster separately in the nucleus), distribution of CpG sites, clustering of compartments or even the clusters of particular promoter/enhancer complexes. The authors of the NucDynamics model paid particular attention to this kind of validation in their original article, though other models have also employed similar validation methods. For example, the authors of Nagano et al.

2013 compared the chromosomal territory diameters of their model with those taken from FISH experiments.

Finally, even without external independent data or artificial ground-truth structures, authors of chromatin modeling methods often resort to other ways to validate their models or at least check whether they work as intended. A popular approach is to check for the number or fraction of contact violations (e.g. [17,37,38,42]) (Fig. 2C). By contact violation, we mean a situation where the distance between a pair of loci connected by a contact is substantially greater than expected based on the potential function. If many violated contacts exist, the model likely fails to converge to a structure that satisfies these restraints. Another method is to calculate the distance matrix from the model structure and compare it either with the original single-cell contact matrix (e.g. [43]) or with bulk Hi-C (e.g. [38]). Very often, the resulting structures are analyzed using some form of clustering analysis (see Fig. 2D). Typical approaches include hierarchical clustering algorithms (e.g. [28,37]) or spectral clustering (e.g. [38]). Among other insights, this type of analysis can provide the authors with the way in which the model can be checked whether it converges to the same or the few preferred conformations, which likely represent local minima of the global potential function applied by the model.

Apart from the problem of model validation, there is also a need for an efficient way of finding the optimal set of model parameters. Given the complexity of chromatin modeling, these models generally involve numerous parameters. The investigation and refinement of these parameters are critical components of molecular simulations, which rely on advanced physical and mathematical theories, such as Hamiltonian systems, statistical physics, and phase transition theory [92–94]. Consequently, three primary approaches to parameter fine-tuning exist (1) direct optimization using experimental datasets, (2) optimization based on a priori biophysical knowledge of observable quantities like the degree of folding or linker length of chromatin, and (3) qualitative comparison of models under varying parameter conditions. Examples of the first approach include dimensionality reduction [37] and heatmap optimization techniques [42,53,61]. Parameters defined by a priori biophysical knowledge are common in nearly all models, such as the Lennard-Jones potential used in [38,43]. Finally, qualitative analysis under parameter variation serves as a powerful tool for understanding polymer structures, as demonstrated in [40], where structural variation of models is evaluated under different values of local variability, smoothness, or prior, and in loop extrusion models [56,57,59,95,96], where authors tested the level of compaction of chromatin under different assumptions for the biophysical laws of loop extrusion. Combining this approach with the other two can yield more reliable models.

4. Discussion

The modeling based on single-cell Hi-C data is very challenging [36,97]. This is partly due to the inherent variability present in the single-cell or single-nucleus data [98]. One of the greatest challenges in modeling chromatin conformation from single-cell Hi-C data is its sparsity and difficulty distinguishing between real contacts and those resulting from spurious random noise [99].

In this review, we described the most common approaches to three-dimensional chromatin structure reconstruction from single-cell Hi-C data. We provided an overview of the field within the broader context of single-cell Hi-C research and averaged bulk Hi-C data modeling. Particular attention was given to how these models structure their objective functions, whether framed in terms of force fields, potentials, cost functions, or likelihood probability functions. Despite appearing to approach the problem from different angles, we found profound general similarities in current chromatin modeling methods.

We also reviewed how chromatin models are validated. A general single-cell Hi-C model can be built from several basic building blocks, as shown schematically in Fig. 1. The two fundamental components include some form of attractive restraint force (usually harmonic) and a

type of repulsive force to account for the excluded volume effect of the modeled polymer. Model creators can add additional terms to their cost functions, such as fluid viscosity force (DPDchrom), variation penalty (SIMBA3D) or conformity to the bulk Hi-C matrix (Nagano et al., 2013), among others. The choice of optimization method—whether molecular dynamics integrators, gradient descent, or simulated annealing—also plays a crucial role. Alternatively, a probabilistic approach could be taken, framing the problem in Bayesian terms using Monte Carlo simulations (ISDHi-C) or a maximum likelihood approach (SIMBA3D). Hierarchical modeling, which uses multiple resolutions and structure interpolation, is often beneficial in achieving the desired structure.

We highlighted the inconsistent approaches to model validation across studied models, emphasizing the need for a comprehensive validation strategy. Despite these challenges, we expect further development of single-cell Hi-C modeling methods. However, due to inherent issues with single-cell experiment analysis, methods relying on bulk Hi-C data are likely to continue growing in number [49,88,100–112].

Validation and comparison of chromatin models are still in their infancy, making it difficult to identify the best models. Some method authors have made little effort to compare their methodologies with previous ones. We described the most common validation ideas in Fig. 2. While the methods' parameters and strategies have likely been optimized in their general outline, it is unclear if they merely represent local minima in the broader landscape of modeling strategies. How accurate would a model be if it combined potential components from different models? How would changing one form of restraint force to another affect overall model behavior? How would models that do not originally use a hierarchical scheme perform when employed in such a manner? The possibilities for chromatin modeling seem endless, with no straightforward way to determine the best approach.

Although chromatin structure reconstruction is the primary focus of this review, it is not the only approach for analyzing sparse interactome data from single-cell Hi-C experiments. Other related fields include interaction imputation, Hi-C map feature calling for single-cell maps, and embedding and clustering sets of single-cell Hi-C data [19]. These methods primarily focus on dimensionality reduction and contact imputation [97]. scHiCluster [113] combines data imputation and single-cell embedding, while methods like Higashi [23] or Fast-Higashi [114] take a comprehensive approach to perform both tasks. New methods, including those using deep generative modeling [115], are constantly being developed. Algorithms using learned embeddings to cluster single-cell data by cell type or cell cycle phase [36] are also emerging. These embeddings, often visualized using techniques like t-SNE [116] or UMAP [117], can form the basis for cell cycle trajectory reconstruction algorithms [118].

4D Nucleome [119], a major consortium in nuclear architecture research, has identified the study of chromatin organization with temporal dynamics as one of its central missions [120]. We believe three-dimensional chromatin models based on single-cell Hi-C data, coupled with inferred cell cycle trajectory information, represent an intriguing direction for future research. Single-cell Hi-C modeling is likely needed to bridge the gap between 3D chromatin modeling and fields such as contact imputation and embedding, where approaches like deep generative modeling play an increasing role [115,121]. In our opinion, models such as those described in this review, if adapted to incorporate contact imputation data and cell cycle trajectory information, could enhance our understanding of chromatin conformation changes during the cell cycle and cell maturation.

While future insights may come from theoretical approaches like MDS, we observe that the field is more likely to develop towards potential optimization-based and physics-based approaches such as molecular dynamics providing mechanistic insights into chromatin folding mechanisms [122]. General molecular dynamics simulation toolkits such as OpenMM [123] or HooMD [81] will likely benefit new method developers and the testing of biological hypotheses related to three-dimensional chromatin conformation. Finally, integrating multiple data modalities

and the broader research area of single-cell embedding, clustering, and cell cycle trajectory inference will likely benefit the field's future development.

5. Funding

Research was funded by Warsaw University of Technology within the Excellence Initiative: Research University (IDUB) programme. This work has been 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”.

CRedit authorship contribution statement

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

Declaration of competing interest

Authors do not acknowledge any conflicts of interest.

References

- Lieberman-Aiden Erez, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009;326(5950):289–93. <https://doi.org/10.1126/science.1181369>. issn: 00368075.
- Fraser James, et al. Chromatin conformation signatures of cellular differentiation. *Genome Biol* 2009;10(4). <https://doi.org/10.1186/gb-2009-10-4-r37>. issn: 14747596.
- Tanizawa Hideki, et al. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res* 2010;38(22):8164–77. <https://doi.org/10.1093/nar/gkq955>. issn: 03051048.
- Rousseau Mathieu, et al. Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinform* 2011;12(1):414. <https://doi.org/10.1186/1471-2105-12-414>.
- Baù Davide, et al. The three-dimensional folding of the α -globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol* 2011;18(1):107–15. <https://doi.org/10.1038/nsmb.1936>. issn: 15459993.
- Noble William, et al. A three-dimensional model of the yeast genome. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. LNBI, vol. 6577. 2011. p. 320. 7296. issn: 16113349.
- Baù Davide, Marti-Renom Marc A. Genome structure determination via 3C-based data integration by the Integrative Modeling Platform. *Methods* 2012;58(3):300–6. <https://doi.org/10.1016/j.ymeth.2012.04.004>. issn: 10462023.
- Gehlen Lutz R, et al. Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus* 2012;3(4):370–83. <https://doi.org/10.4161/nucl.20971>. issn: 19491042.
- Reza Kalhor, et al. Solid-phase chromosome conformation capture for structural characterization of genome architectures. *Nat Biotechnol* 2012;30(1):90–8. <https://doi.org/10.1038/nbt.2057>. issn: 15378276.
- Hu Ming, et al. Bayesian inference of spatial organizations of chromosomes. *PLoS Comput Biol* 2013;9(1). <https://doi.org/10.1371/journal.pcbi.1002893>. issn: 15537358.
- Meluzzi Dario, Arya Gaurav. Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Res* 2013;41(1):63–75. <https://doi.org/10.1093/nar/gks1029>. issn: 03051048.
- Arsuaga Javier, et al. Current theoretical models fail to predict the topological complexity of the human genome. *Front Mol Biosci* 2015;2. <https://doi.org/10.3389/fmolb.2015.00048>. issn: 2296889X.
- Wettermann S, et al. A minimal G₀-model for rebuilding whole genome structures from haploid single-cell Hi-C data. *Comput Mater Sci* 2020;173. <https://doi.org/10.1016/j.commatsci.2019.109178>. issn: 09270256.
- Rothörl Jan, et al. Reconstructing diploid 3D chromatin structures from single cell Hi-C data with a polymer-based approach. *Front Bioinform* 2023;3:1–8. <https://doi.org/10.3389/fbinf.2023.1284484>. issn: 26737647.
- Lajoie Bryan R, Dekker Job, Kaplan Noam. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* 2015;72:65–75. <https://doi.org/10.1016/j.ymeth.2014.10.031>. issn: 10959130.
- Segal Mark R. Can 3D diploid genome reconstruction from unphased Hi-C data be salvaged? *NAR Genomics Bioinform* 2022;4. <https://doi.org/10.1093/nargab/lqac038>. issn: 26319268.
- Nagano Takashi, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 2013;502(7469):59–64. <https://doi.org/10.1038/nature12593>. issn: 00280836.
- Gassler Johanna, et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J* 2017;36(24):3600–18. <https://doi.org/10.15252/embj.201798083>. issn: 0261-4189.
- Galitsyna Aleksandra A, Gelfand Mikhail S. Single-cell Hi-C data analysis: safety in numbers. *Brief Bioinform* 2021;22(6):1–13. <https://doi.org/10.1093/bib/bbab316>. issn: 14774054.
- Gong Haiyan, Ma Fuqiang, Zhang Xiaotong. Advances in methods and applications of single-cell Hi-C data analysis. *J Biomed Eng* 2023;40(5):1033–9. <https://doi.org/10.7507/1001-5515.202303046>. issn: 1001-5515.
- Fudenberg Geoff, Kelley David R, Pollard Katherine S. Predicting 3D genome folding from DNA sequence with Akita. *Nat Methods* 2020;17(11):1111–7. <https://doi.org/10.1038/s41592-020-0958-x>.
- Yang Rui, et al. Epiphany: predicting Hi-C contact maps from 1D epigenomic signals. *Genome Biol* 2023;24(1). <https://doi.org/10.1186/s13059-023-02934-9>.
- Zhang Ruochi, Zhou Tianming, Ma Jian. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat Biotechnol* 2022;40(2):254–61. <https://doi.org/10.1038/s41587-021-01034-y>. issn: 15461696.
- Valeyre Henry, et al. CHROMFORMER: a transformer-based model for 3D genome structure prediction. <https://doi.org/10.1101/2022.11.15.516571>. <https://github.com/AI4SCR/ChromFormer>, 2022.
- Lesne Annick, et al. 3D genome reconstruction from chromosomal contacts. *Nat Methods* 2014;11(11):1141–3. <https://doi.org/10.1038/nmeth.3104>. issn: 15487105.
- Li Jiangeng, Zhang Wei, Li Xiaodan. 3D genome reconstruction with ShRec3D+ and Hi-C data. *IEEE/ACM Trans Comput Biol Bioinform* 2018;15(2):460–8. <https://doi.org/10.1109/TCBB.2016.2535372>. issn: 15455963.
- Nagano Takashi, et al. Cell-cycle dynamics of chromosomal organisation at single-cell resolution. *Nature* 2017;547(7661):61–7. <https://doi.org/10.1038/nature23001>. issn: 14764687.
- Stevens Tim J, et al. 3D structure of individual mammalian genomes studied by single cell Hi-C. *Nature* 2017;544(7648):59–64. <https://doi.org/10.1038/nature21429>.
- Bonev Boyan, et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* 2017;171(3):557–57224.e. <https://doi.org/10.1016/j.cell.2017.09.043>. issn: 0092-8674.
- Tan Longzhi, et al. Three-dimensional genome structures of single diploid human cells. *Science* 2018;361(6405):924–8. <https://doi.org/10.1126/science.aat5641>. issn: 10959203.
- Tan Longzhi, et al. Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. *Nat Struct Mol Biol* 2019;26:297–307. <https://doi.org/10.1038/s41594-019-0205-2>. doi.
- Flyamer Ilya M, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 2017;544(7648):110–4. <https://doi.org/10.1038/nature21711>. issn: 14764687.
- Ramani Vijay, et al. Massively multiplex single-cell Hi-C. *Nat Methods* 2017;14(3):263–6. <https://doi.org/10.1038/nmeth.4155>. issn: 15487105.
- Lee Dong Sung, et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat Methods* 2019;16(10):999–1006. <https://doi.org/10.1038/s41592-019-0547-z>. issn: 15487105.
- Collombet Samuel, et al. Parental-to-embryo switch of chromosome organization in early embryogenesis. *Nature* 2020;580(7801):142–6. <https://doi.org/10.1038/s41586-020-2125-z>. issn: 14764687.
- Zhen Caiwei, et al. A review and performance evaluation of clustering frameworks for single-cell Hi-C data. <https://doi.org/10.1093/bib/bbac385>, Nov. 2022.
- Paulsen Jonas, Gramstad Odin, Collas Philippe. Manifold based optimization for single-cell 3D genome reconstruction. *PLoS Comput Biol* 2015;11(8). <https://doi.org/10.1371/journal.pcbi.1004396>. issn: 15537358.
- Carstens Simeon, Nilges Michael, Habeck Michael. Inferential structure determination of chromosomes from single-cell Hi-C data. *PLoS Comput Biol* 2016;12:1–33. <https://doi.org/10.1371/journal.pcbi.1005292>.
- Hirata Yoshito, et al. Three-dimensional reconstruction of single-cell chromosome structure using recurrence plots. *Sci Rep* 2016;6:3–8. <https://doi.org/10.1038/srep34982>. issn: 20452322.
- Rosenthal Michael, et al. Bayesian estimation of three-dimensional chromosomal structure from single-cell Hi-C data. *J Comput Biol* 2019;26(11):1191–202. <https://doi.org/10.1089/cmb.2019.0100>. issn: 10665277.

- [41] Zhu Hao, Wang Zheng. SCL: A lattice-based approach to infer 3D chromosome structures from single-cell Hi-C data. *Bioinformatics* 2019;35(20):3981–8. <https://doi.org/10.1093/bioinformatics/btz181>. issn: 14602059.
- [42] Meng Luming, et al. Si-C is a method for inferring super-resolution intact genome structure from single-cell Hi-C data. *Nat Commun* 2021;12(1):1–11. <https://doi.org/10.1038/s41467-021-24662-z>. issn: 20411723.
- [43] Zha Mengsheng, et al. Inferring single-cell 3d chromosomal structures based on the Lennard-Jones potential. *Int J Mol Sci* 2021;22(11). <https://doi.org/10.3390/ijms22115914>. issn: 14220067.
- [44] Kos Pavel I, et al. Perspectives for the reconstruction of 3D chromatin conformation using single cell Hi-C data. *PLoS Comput Biol* 2021;17(11). <https://doi.org/10.1371/journal.pcbi.1009546>. issn: 15537358.
- [45] Fudenberg Geoffrey, et al. Formation of chromosomal domains by loop extrusion. *Cell Rep* 2016;15(9):2038–49. <https://doi.org/10.1016/j.celrep.2016.04.085>. issn: 22111247.
- [46] Oluwadare Oluwatosin, Highsmith Max, Cheng Jianlin. An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data. *Biol Proced Online* 2019;21(1):1–20. <https://doi.org/10.1186/s12575-019-0094-0>. issn: 14809222.
- [47] MacKay Kimberly, Kusalik Anthony. Computational methods for predicting 3D genomic organization from high-resolution chromosome conformation capture data. *Brief Funct Genomics* 2020;19(4):292–308. <https://doi.org/10.1093/bfgp/ela004>. issn: 20412657.
- [48] Varoquaux Nelle, et al. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics* 2014;30(12):i26–33. <https://doi.org/10.1093/bioinformatics/btu268>. issn: 14602059.
- [49] Wang Hao, et al. Reconstruct high-resolution 3D genome structures for diverse cell-types using FLAMINGO. *Nat Commun* 2022;13(1):1–18. <https://doi.org/10.1038/s41467-022-30270-2>. issn: 20411723.
- [50] Zhang ZhiZhao, et al. 3D chromosome modeling with semi-definite programming and Hi-C data. *J Comput Biol* 2013;20(11):831–46. <https://doi.org/10.1089/cmb.2013.0076>. PMID: 24195706.
- [51] Trieu Tuan, Cheng Jianlin. 3D genome structure modeling by Lorentzian objective function. *Nucleic Acids Res* 2017;45(3):1049–58. <https://doi.org/10.1093/nar/gkw1155>. issn: 13624962.
- [52] Fudenberg Geoffrey, Mirny Leonid A. Higher-order chromatin structure: bridging physics and biology. <https://doi.org/10.1016/j.gde.2012.01.006>, Apr. 2012.
- [53] Ay Ferhat, et al. Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res* 2014;24(6):974–88. <https://doi.org/10.1101/gr.169417.113>. issn: 15495469.
- [54] Imakaev Maxim V, Fudenberg Geoffrey, Mirny Leonid A. Modeling chromosomes: beyond pretty pictures. <https://doi.org/10.1016/j.febslet.2015.09.004>, Oct. 2015.
- [55] Fudenberg Geoffrey, et al. Emerging evidence of chromosome folding by loop extrusion. *Cold Spring Harbor symposia on quantitative biology*, vol. 82. Cold Spring Harbor Laboratory Press; 2017. p. 45–55.
- [56] Goloborodko Anton, Marko John F, Mirny Leonid A. Chromosome compaction by active loop extrusion. *Biophys J* 2016;110(10):2162–8. <https://doi.org/10.1016/j.bpj.2016.02.041>.
- [57] Korsak Sevastianos, Plewczynski Dariusz. LoopSage: an energy-based Monte Carlo approach for the loop extrusion modeling of chromatin. *Methods* 2024;223:106–17. <https://doi.org/10.1016/j.ymeth.2024.01.015>. doi.
- [58] Rossini Roberto, et al. MoDLE: high-performance stochastic modeling of DNA loop extrusion interactions. *Genome Biol* 2022;23(1):247. <https://doi.org/10.1186/s13059-022-02815-7>.
- [59] Banigan Edward J, Mirny Leonid A. 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>. *Cell Nucleus*. issn: 0955-0674.
- [60] Beckwith KS, et al. Visualization of loop extrusion by DNA nanoscale tracing in single human cells. *BioRxiv* 2021. pp. 2021–2024.
- [61] Liu Lei, Kim Min Hyeok, Hyeon Changbong. Heterogeneous loop model to infer 3D chromosome structures from Hi-C. *Biophys J* 2019;117(3):613–25. <https://doi.org/10.1016/j.bpj.2019.06.032>. issn: 15420086.
- [62] Zou Chenchen, Zhang Yuping, Ouyang Zhengqing. HSA: integrating multi-track Hi-C data for genome-scale reconstruction of 3D chromatin structure. *Genome Biol* 2016;17(1):1–14. <https://doi.org/10.1186/s13059-016-0896-1>. issn: 1474760X.
- [63] Szalaj Przemyslaw, et al. 3D-GNOME: an integrated web service for structural modeling of the 3D genome. *Nucleic Acids Res* 2016;44(W1):W288–93. <https://doi.org/10.1093/NAR/GKW437>. issn: 13624962.
- [64] Hua Kang Jian, Ma Bin Guang. EVR: reconstruction of bacterial chromosome 3D structure models using error-vector resultant algorithm. *BMC Genomics* 2019;20(1):1–10. <https://doi.org/10.1186/s12864-019-6096-0>. issn: 14712164.
- [65] Abbas Ahmed, et al. Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nat Commun* 2019;10(1). <https://doi.org/10.1038/s41467-019-10005-6>. issn: 20411723.
- [66] Trieu Tuan, Cheng Jianlin. Large-scale reconstruction of 3D structures of human chromosomes from chromosomal contact data. *Nucleic Acids Res* 2014;42(7):1–11. <https://doi.org/10.1093/nar/gkt1411>. issn: 13624962.
- [67] Adhikari Badri, Trieu Tuan, Cheng Jianlin. Chromosome3D: reconstructing three-dimensional chromosomal structures from Hi-C interaction frequency data using distance geometry simulated annealing. *BMC Genomics* 2016;17(1):1–9. <https://doi.org/10.1186/s12864-016-3210-4>. issn: 14712164.
- [68] Zhu Guangxiang, et al. Reconstructing spatial organizations of chromosomes through manifold learning. *Nucleic Acids Res* 2018;46(8):E50. <https://doi.org/10.1093/NAR/GKY065>. issn: 13624962.
- [69] Trieu Tuan, Oluwadare Oluwatosin, Cheng Jianlin. Hierarchical reconstruction of high-resolution 3D models of large chromosomes. *Sci Rep* 2019;9(1):1–12. <https://doi.org/10.1038/s41598-019-41369-w>. issn: 20452322.
- [70] Giorgetti Luca, et al. Chromosome conformation and transcription. *Cell* 2014;157(4):950–63. <https://doi.org/10.1016/j.cell.2014.03.025>. Predictive.
- [71] Tjong Harianto, et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc Natl Acad Sci USA* 2016;113(12):E1663–72. <https://doi.org/10.1073/pnas.1512577113>. issn: 10916490.
- [72] Sefer Emre, Duggal Geet, Kingsford Carl. Deconvolution of ensemble chromatin interaction data reveals the latent mixing structures in cell subpopulations. *J Comput Biol* 2016;23(6):425–38. <https://doi.org/10.1089/cmb.2015.0210>. issn: 10665277.
- [73] Kibble Tom, Berkshire Frank H. *Classical mechanics*. World Scientific Publishing Company; 2004.
- [74] Lenhard Johannes, Stephan Simon, Hasse Hans. On the history of the Lennard-Jones potential. *Ann Phys* 2024;2400115.
- [75] Wang Xipeng, et al. The Lennard-Jones potential: when (not) to use it. *Phys Chem Chem Phys* 2020;22(19):10624–33.
- [76] Erdel Fabian. Biophysical mechanisms of chromatin patterning. *Curr Opin Genet Dev* 2020;61:62–8. <https://doi.org/10.1016/j.gde.2020.03.006>.
- [77] Berne Bruce J, Pechukas Philip. Gaussian model potentials for molecular interactions. *J Chem Phys* 1972;56(8):4213–6. <https://doi.org/10.1063/1.1677837>. doi.
- [78] Jost Daniel, et al. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res* 2014;42(15):9553–61. <https://doi.org/10.1093/nar/gku698>.
- [79] Zhou Rui, Gao Yi Qin. Polymer models for the mechanisms of chromatin 3D folding: review and perspective. *Phys Chem Chem Phys* 2020;22(36):20189–201. <https://doi.org/10.1039/d0cp01877e>.
- [80] Taketomi Hiroshi, Ueda Yuzo, Gō Nobuhiro. Studies on protein folding, unfolding and fluctuations by computer simulation. *Int J Pept Protein Res* 1975;7(6):445–59. <https://doi.org/10.1111/j.1399-3011.1975.tb02465.x>. issn: 13993011.
- [81] Anderson Joshua A, Lorenz Chris D, Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J Comput Phys* 2008;227(10):5342–59. <https://doi.org/10.1016/j.jcp.2008.01.047>. issn: 10902716.
- [82] Glaser Jens, et al. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Comput Phys Commun* 2015;192:97–107. <https://doi.org/10.1016/j.cpc.2015.02.028>. issn: 00104655.
- [83] Oluwadare Oluwatosin, Zhang Yuxiang, Cheng Jianlin. A maximum likelihood algorithm for reconstructing 3D structures of human chromosomes from chromosomal contact data. *BMC Genomics* 2018;19(1):1–17. <https://doi.org/10.1186/s12864-018-4546-8>. issn: 14712164.
- [84] Marwan Norbert, et al. Recurrence plots for the analysis of complex systems. <https://doi.org/10.1016/j.physrep.2006.11.001>, Jan. 2007.
- [85] Kadlof Michal, Rozycka Julia, Plewczynski Dariusz. Spring model – chromatin modeling tool based on OpenMM. *Methods* 2020;181–182:62–9. <https://doi.org/10.1016/j.ymeth.2019.11.014>. issn: 10959130.
- [86] Plewczynski Dariusz, Kadlof Michal. Computational modelling of three-dimensional genome structure. *Methods* 2020;181–182:1–4. <https://doi.org/10.1016/j.ymeth.2020.09.013>. issn: 1046-2023.
- [87] Maris Dimitris T, Goussis Dimitris A. The “hidden” dynamics of the Rössler attractor. *Phys D: Nonlinear Phenom* 2015;295:66–90. <https://doi.org/10.1016/j.physd.2014.12.010>. doi.
- [88] Carstens Simeon, Nilges Michael, Habeck Michael. Bayesian inference of chromatin structure ensembles from population-averaged contact data. *Proc Natl Acad Sci USA* 2020;117(14):7824–30. <https://doi.org/10.1073/pnas.1910364117>. issn: 10916490.
- [89] Beagrie Robert A, et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 2017;543(7646):519–24. <https://doi.org/10.1038/nature21411>. issn: 14764687.
- [90] Bogdan Bintu, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Physiol Behav* 2018;362(6413):139–48. <https://doi.org/10.1126/science.aau1783>.
- [91] Su Jun Han, et al. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* 2020;182(6):1641–59. <https://doi.org/10.1016/j.cell.2020.07.032>. issn: 10974172.
- [92] Tuckerman Mark E, Martyna Glenn J. Understanding modern molecular dynamics: techniques and applications. *J Phys Chem B* 2001;31(105):7598. <https://doi.org/10.1021/jp992433y>.
- [93] Khokhlov Alexei R, Grosberg Alexander Yu, Pande Vijay S. *Statistical physics of macromolecules*, vol. 1. Springer; 1994.
- [94] Strogatz Steven H. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC Press; 2018.
- [95] Banigan Edward J, et al. Chromosome organization by one-sided and two-sided loop extrusion. *eLife* 2020;9:e53558. <https://doi.org/10.7554/eLife.53558>.

- [96] Banigan Edward J, Mirny Leonid A. The interplay between asymmetric and symmetric DNA loop extrusion. *eLife* 2020;9:e63528. <https://doi.org/10.7554/eLife.63528>.
- [97] Zhang Yang, et al. Computational methods for analysing multiscale 3D genome organization. *Nat Rev Genet* 2024;25(2):123–41. <https://doi.org/10.1038/s41576-023-00638-1>. issn: 14710064.
- [98] Finn Elizabeth H, Misteli Tom. Molecular basis and biological function of variability in spatial genome organization. <https://doi.org/10.1126/science.aaw9498>, Sept. 2019.
- [99] Sekelja Monika, Paulsen Jonas, Collas Philippe. 4D nucleomes in single cells: what can computational modeling reveal about spatial chromatin conformation? *Genome Biol* 2016;17(1):2. <https://doi.org/10.1186/s13059-016-0923-2>. issn: 1474760X.
- [100] Zhen Li Fang, et al. Chromatin 3D structure reconstruction with consideration of adjacency relationship among genomic loci. *BMC Bioinform* 2020;21(1):1–17. <https://doi.org/10.1186/s12859-020-03612-4>. issn: 14712105.
- [101] Shinkai Soya, et al. PHi-C: deciphering Hi-C data into polymer dynamics. *NAR Genomics Bioinform* 2020;2(2):1–10. <https://doi.org/10.1093/nargab/lqaa020>. issn: 26319268.
- [102] Zhang Rongrong, et al. Inferring spatial organization of individual topologically associated domains via piecewise helical model. *IEEE/ACM Trans Comput Biol Bioinform* 2020;17(2):647–56. <https://doi.org/10.1109/TCBB.2018.2865349>. Inferring.
- [103] Perez-Rathke Alan, et al. CHROMATIX: computing the functional landscape of many-body chromatin interactions in transcriptionally active loci from deconvolved single cells. *Genome Biol Jan.* 2020;21(1). <https://doi.org/10.1186/s13059-019-1904-z>. issn: 1474760X.
- [104] Liang Jie, Perez-Rathke Alan. Minimalistic 3D chromatin models: sparse interactions in single cells drive the chromatin fold and form many-body units. *Curr Opin Struct Biol* 2021;71:200–14. <https://doi.org/10.1016/j.sbi.2021.06.017>. issn: 1879033X.
- [105] Collins Brandon, Oluwadare Oluwatosin, Brown Philip. Chromebat: a bio-inspired approach to 3d genome reconstruction. *Genes* 2021;12(1). <https://doi.org/10.3390/genes12111757>. issn: 20734425.
- [106] Lappala Anna, et al. Four-dimensional chromosome reconstruction elucidates the spatiotemporal reorganization of the mammalian X chromosome. *Proc Natl Acad Sci USA* 2021;118(42). <https://doi.org/10.1073/pnas.2107092118>. issn: 10916490.
- [107] Messelink Joris JB, et al. Learning the distribution of single-cell chromosome conformations in bacteria reveals emergent order across genomic scales. *Nat Commun* 2021;12(1). <https://doi.org/10.1038/s41467-021-22189-x>. issn: 20411723.
- [108] Wasim Abdul, Gupta Ankit, Mondal Jagannath. A Hi-C data-integrated model elucidates E. coli chromosome's multiscale organization at various replication stages. *Nucleic Acids Res* 2021;49(6):3077–91. <https://doi.org/10.1093/nar/gkab094>. issn: 13624962.
- [109] Shinkai Soya, et al. PHi-C2: interpreting Hi-C data as the dynamic 3D genome state. *Bioinformatics* 2022;38(21):4984–6. <https://doi.org/10.1093/bioinformatics/btac613>. issn: 13674811.
- [110] Tuzhilina Elena, Hastie Trevor J, Segal Mark R. Principal curve approaches for inferring 3D chromatin architecture. *Biostatistics* 2022;23(2):626–42. <https://doi.org/10.1093/biostatistics/kxaa046>. issn: 14684357.
- [111] Shi Guang, 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). <https://doi.org/10.1038/s41467-023-36412-4>. issn: 20411723.
- [112] Li Zilong, Schlick Tamar. Hi-BDiSCO: folding 3D mesoscale genome structures from Hi-C data using Brownian dynamics. *Nucleic Acids Res* 2024;52(2):583–99. <https://doi.org/10.1093/nar/gkad1121>. issn: 13624962.
- [113] Zhou Jingtian, et al. Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc Natl Acad Sci USA* 2019;116(28):14011–8. <https://doi.org/10.1073/pnas.1901423116>. issn: 10916490.
- [114] Zhang Ruochi, Zhou Tianming, Ma Jian. Ultrafast and interpretable single-cell 3D genome analysis with Fast-Higashi. *Cell Syst Oct.* 2022;13(10):798–807. <https://doi.org/10.1016/j.cels.2022.09.004>. issn: 24054720.
- [115] Liu Qiao, et al. Deep generative modeling and clustering of single cell Hi-C data. *Brief Bioinform* 2023;24(1). <https://doi.org/10.1093/bib/bbac494>. issn: 14774054.
- [116] van der Maaten Laurens, Hinton Geoffrey. Visualizing data using t-SNE. *J Mach Learn Res* 2008;9:2579–605.
- [117] McInnes Leland, et al. UMAP: uniform manifold approximation and projection. Preprint. arXiv:1802.03426, 2018.
- [118] Ye Yusen, Gao Lin, Zhang Shihua. Circular trajectory reconstruction uncovers cell-cycle progression and regulatory dynamics from single-cell Hi-C maps. *Adv Sci* 2019;6(23). <https://doi.org/10.1002/advs.201900986>. issn: 21983844.
- [119] Dekker Job, et al. The 4D nucleome project. <https://doi.org/10.1038/nature23884>, Sept. 2017.
- [120] Dekker Job, et al. Spatial and temporal organization of the genome: current state and future aims of the 4D nucleome project. <https://doi.org/10.1016/j.molcel.2023.06.018>, Aug. 2023.
- [121] Zheng Ye, Shen Siqi, Keleş Sündüz. Normalization and de-noising of single-cell Hi-C data with BandNorm and scVI-3D. *Genome Biol* 2022;23(1). <https://doi.org/10.1186/s13059-022-02774-z>. issn: 1474760X.
- [122] Portillo-Ledesma Stephanie, Li Zilong, Schlick Tamar. Genome modeling: from chromatin fibers to genes. <https://doi.org/10.1016/j.sbi.2022.102506>, Feb. 2023.
- [123] Eastman Peter, et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* 2017;13(7). <https://doi.org/10.1371/journal.pcbi.1005659>. issn: 15537358.