



Article

Inferring Single-Cell 3D Chromosomal Structures Based on the Lennard-Jones Potential

Mengsheng Zha ¹, Nan Wang ², Chaoyang Zhang ¹ and Zheng Wang ^{3,*}

¹ School of Computing Sciences and Computer Engineering, University of Southern Mississippi, 118 College Dr, Hattiesburg, MS 39406, USA; mengsheng.zha@usm.edu (M.Z.); chaoyang.zhang@usm.edu (C.Z.)

² Department of Computer Science, New Jersey City University, 2039 Kennedy Blvd, Jersey City, NJ 07305, USA; nwang1@njcu.edu

³ Department of Computer Science, University of Miami, 1364 Memorial Drive, Coral Gables, FL 33124, USA

* Correspondence: zheng.wang@miami.edu

Abstract: Reconstructing three-dimensional (3D) chromosomal structures based on single-cell Hi-C data is a challenging scientific problem due to the extreme sparseness of the single-cell Hi-C data. In this research, we used the Lennard-Jones potential to reconstruct both 500 kb and high-resolution 50 kb chromosomal structures based on single-cell Hi-C data. A chromosome was represented by a string of 500 kb or 50 kb DNA beads and put into a 3D cubic lattice for simulations. A 2D Gaussian function was used to impute the sparse single-cell Hi-C contact matrices. We designed a novel loss function based on the Lennard-Jones potential, in which the ϵ value, i.e., the well depth, was used to indicate how stable the binding of every pair of beads is. For the bead pairs that have single-cell Hi-C contacts and their neighboring bead pairs, the loss function assigns them stronger binding stability. The Metropolis–Hastings algorithm was used to try different locations for the DNA beads, and simulated annealing was used to optimize the loss function. We proved the correctness and validity of the reconstructed 3D structures by evaluating the models according to multiple criteria and comparing the models with 3D-FISH data.

Keywords: 3D genome; single-cell Hi-C; 3D chromosomal structure; Lennard-Jones potential



Citation: Zha, M.; Wang, N.; Zhang, C.; Wang, Z. Inferring Single-Cell 3D Chromosomal Structures Based on the Lennard-Jones Potential. *Int. J. Mol. Sci.* **2021**, *22*, 5914. <https://doi.org/10.3390/ijms22115914>

Academic Editors: Daisuke Kihara and Jianlin Cheng

Received: 22 April 2021

Accepted: 28 May 2021

Published: 31 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Compared to the study of the genome from a one-dimensional (1D) or sequence perspective, the research of spatial folding of the 3D structure of the DNA has been increasingly recognized as important [1–3] because the spatial locations of genes, enhancers, promoters, and lncRNAs are essential for biological functions such as transcription, replication, DNA repair, and chromosome translocation [4,5]. Based on the spatial folding of the DNA, researchers have defined on average 1 Mb long topologically associating domains (TADs) on the genome [6], which can be the structural and functional units of the genome. Research such as the reconstruction of high-resolution 3D genome structures [7] and family classification of TADs [8] lead to a deeper understanding of how the genome functions.

Empowered by the methods of chromosome conformation capture including 3C, 4C, 5C, and Hi-C, the technology of genome conformation detection is applicable to not only a selection of loci of chromosomes but also higher-order structures such as the entire genome. Specifically, 3C methods can detect the proximate or in-contact relationship in the three-dimensional (3D) space between a single location in the chromosome and other selected genomic regions (single vs. selected) [9–18]; 4C methods can detect the same in-contact relationship between one genomic region and all other genomic regions (one vs. all) [19–25]; 5C can detect the spatial proximities between selective genomic regions and other selective genomic regions (many vs many) [26–31]; and Hi-C, as one of the latest chromosome conformation capturing methods, can detect genome-wide all-to-all spatial in-contact relationships [32–54].

Each cell contains a complete genome, which consists of multiple chromosomes. Each chromosome is a two-strand (+ and −) string of nucleotides (A, G, C, and T). The Hi-C experiment can detect which two locations in the same chromosome (intra-chromosomal) or two different chromosomes (inter-chromosomal) are spatially proximate or in contact in the 3D space. Usually, the entire chromosome is represented using a beads-on-a-string approach, in which a certain number of nucleotides, for example, 500 kb (kilo DNA base pairs), is considered a bead, and the chromosome is considered a long string of 500 kb DNA beads. The 500 kb length for each bead is also called the resolution of the 3D chromosome structure. The lower the length of each bead is, the higher resolution the reconstructed 3D structure is at, and a higher resolution leads to a higher number of beads for the same chromosome.

Using the Hi-C method as an example, the intra- and inter-chromosomal contact maps can be generated to indicate the number of Hi-C contacts detected between two regions of the same or different chromosome(s). Based on a Hi-C contact map, the 3D structures of the DNA can be reconstructed. The 3D chromosomal structures of yeast were reconstructed to find relationships between the function and structure of the eukaryotic genome [55]. The computational module named Integrative Modeling Platform (IMP) can be used to measure and evaluate the 3D structure of genomic regions and the whole genome. Specifically, [56] used IMP to model 3D chromosomal structures, from which the authors found the characteristics of each chromatin and their relationships to gene expression. The interaction frequencies (IF) detected by 5C and Hi-C have both been used to computationally model and analyze three-dimensional chromatin organization in [57]. Data noise and structural changeability were used to evaluate the mean-field restraint-based methods of 3D genome structure reconstruction and were used to verify the limitations of the 3D modeling methods [58]. The semi-definite programming techniques have been applied in ChromSDE, a method to figure out the best structure that matches the observed data and discover the optimal parameters to transform Hi-C contact numbers to Euclidean distance in the 3D space using golden section search [59]. BACH, a Bayesian 3D genome structure reconstructing algorithm based on Hi-C data, together with its updated version BACH-MIX have been used to search the structural variations of chromatin in a cell population and find the consensus 3D chromosomal structure. Based on the experiments of applying BACH and BACH-MIX to mouse embryonic stem cells, BACH and BACH-MIX were considered to have the potential to accurately reconstruct the chromosomal structures of mammalian cells [60]. [61] claimed that Poisson models can better reconstruct chromosomal structures in comparison to multidimensional scaling (MDS)-based methods since Poisson models produce more reproducible structures than MDS-based methods. Last but not the least, the 3D chromosomal reconstruction method named MOGEN uses a data-driven scoring function to model the in-contact and not-in-contact relationships between pairs of DNA segments in the same or different chromosome(s) [62]. miniMDS [63] builds 10 kb resolution chromosomal structures by partitioning a Hi-C dataset, performing high-resolution MDS on each partition, and reassembling the inferred structure for each partition using low-resolution MDS. 3D-GNOME 2.0 [64] can predict the alternations of chromosomal 3D structures driven by structural variants given the input of Hi-C data, ChIA-PET data, CTCF interactions, or CTCF and RNPII interactions. GEM [65] reconstructs 3D chromosomal structures by integrating Hi-C data with biophysical feasibility and directly embeds the neighboring affinities from the Hi-C space into the 3D Euclidean space. GEM-FISH [66] uses both Hi-C and FISH data and a polymer biophysical model to infer 3D chromosomal structures and reveals spatial distribution patterns of super-enhancers. ChromStruct4 [67] is a multiresolution method that uses modified-bead-chain representation of chromatin and Monte Carlo sampling to build a set of 3D configurations that fit the Hi-C data and prior knowledge. PGS [68] can deconvolve the population-cell Hi-C data and generate a population of distinct diploid 3D genome structures that match inter-chromosomal contact patterns.

The Hi-C method mentioned above is a biochemistry experiment conducted on a population of cells, usually in the range of 10^5 cells. Therefore, it is often called population-cell Hi-C or bulk Hi-C. Because each cell has a complete genome, the population Hi-C contacts are a consensus dataset over a pool of cells. In other words, millions of Hi-C contacts can usually be generated. Therefore, the common idea of the population-Hi-C-based methods for reconstructing 3D chromosome structure is that a higher number of contacts detected between a pair of DNA beads indicates a shorter Euclidean distance between the two beads in the reconstructed 3D structure.

The recent development of biochemistry technology began to focus on individual cells or single cells because that allows scientists to understand the accurate knowledge about every single cell instead of a consensus result from a population of cells. The same trend has also happened in the field of chromosome conformation capturing. Recently, single-cell Hi-C techniques have been invented and applied to mouse and human cells [69–74]. The bioinformatics community would naturally want to design algorithms to model the 3D structure of a single cell because of the potential significance and novelty of the research. However, modeling 3D chromosomal structures based on single-cell Hi-C data is a difficult computational task because the number of Hi-C contacts observed from a single-cell Hi-C experiment is usually much less than the ones generated from a population-cell Hi-C experiment. In other words, zero is the dominant value in the single-cell Hi-C contact matrices, making the matrices highly sparse. [75] used a modified Bayesian inferential structure determination (ISD) framework to address this challenge. [69] modeled the 3D chromosomal structures based on single-cell Hi-C data using molecular dynamics and simulated annealing. SIMBA3D [76] is a Bayesian multiscale approach for inferring single-cell chromosomal structures using population-cell Hi-C data as prior. None of these methods considered the influence that a single-cell Hi-C contact can generate on the sequentially neighboring bead pairs. SCL [77] uses a 2D Gaussian function to impute the influence that a single-cell Hi-C contact generates to its sequentially neighboring bead pairs. In this research, we adapted the same 2D Gaussian imputation approach but used the Lennard-Jones potential to model the single-cell 3D chromosomal structures, which to the best of our knowledge is the first attempt in the field.

Polymer physics has been applied to model and analyze chromatin structures. In [78], the authors modeled the structure of TADs based on loop extrusion and found that the cis-acting loop-extruding factors, likely cohesins, form large loops but stall at the TAD boundaries. Also based on loop extrusion and using only information about the binding locations of CTCF, the authors of [79] accurately predicted how the genome will fold. The “strings and binder switch” (SBS) model can model the chromatin folding changes based on the binding site distribution, binder concentration, or binding affinity in a switch-like fashion, which has been used in [80] to capture the complexity of chromatin folding. The authors of [81] at first used machine learning to predict binding locations based on population-cell Hi-C data and then also used the phrase-separation SBS model to indicate chromatin folding variability across single cells. Other studies that also used phrase-separation models to build and analyze 3D DNA structures include [80,82,83].

The Lennard-Jones potential has been widely used for molecular dynamics simulations [84–88] as its mathematical formula can very closely mimic the real repulsive and attractive forces between molecules. [89] used the Lennard-Jones potential to model the 3D structures of proteins. In [82], every 10 kb of DNA is represented as a monomer, and a truncated Lennard-Jones-like potential is used to model the energy between monomers. The authors of [79] used the Lennard-Jones potential to model both the intermonomer attractions and external crushing forces during condensation when they used loop extrusion to model the 3D structures of wild-type and engineered genomes. However, the Lennard-Jones potential has not been used to model genome 3D structures based on extremely sparse single-cell Hi-C data.

2. Results and Evaluations

2.1. Inferred 3D Structures of the X-Chromosome of a Mouse TH1 Cell

We downloaded the single-cell Hi-C data of individual male mouse TH1 cells from [69], in which high-quality single-cell Hi-C data were generated on 10 cells. The single-cell Hi-C experiment on cell 1 has the highest quality based on [69]. Therefore, it was used in this research. There are 616 single-cell intra-X-chromosome Hi-C contacts in the dataset. After removing self-contacts, 438 single-cell Hi-C contacts were kept and used as the input of our algorithm.

Figure 1a shows the 3D structure of the X-chromosome at 50 kb resolution inferred by our algorithm. In Figure 1b, each black dot represents a single-cell Hi-C contact detected for the Th1 X-chromosome. The heatmap overlapped with the single-cell Hi-C contacts represents the Euclidean distance of every bead pair that is parsed from the inferred 3D structure. Lighter orange color represents shorter distances, whereas darker orange color represents longer distances. Therefore, if we see that white or lighter colors are mostly overlapped with black dots (single-cell Hi-C contacts), that means the reconstructed 3D structure makes sense or fits the single-cell Hi-C data. From Figure 1b, we can see that all of the reconstructed 3D structures contain large regions that well fit the single-cell Hi-C data.

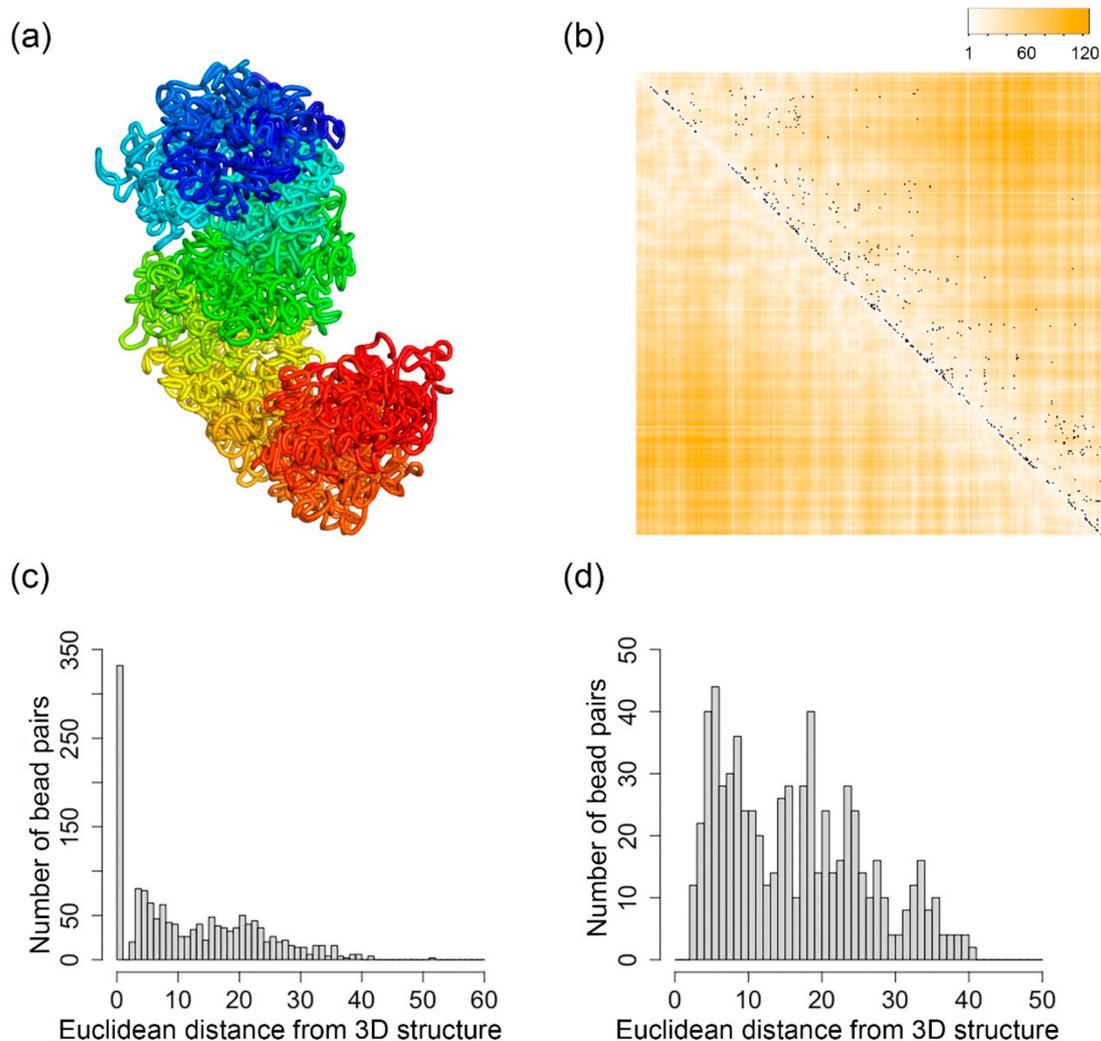


Figure 1. Cont.

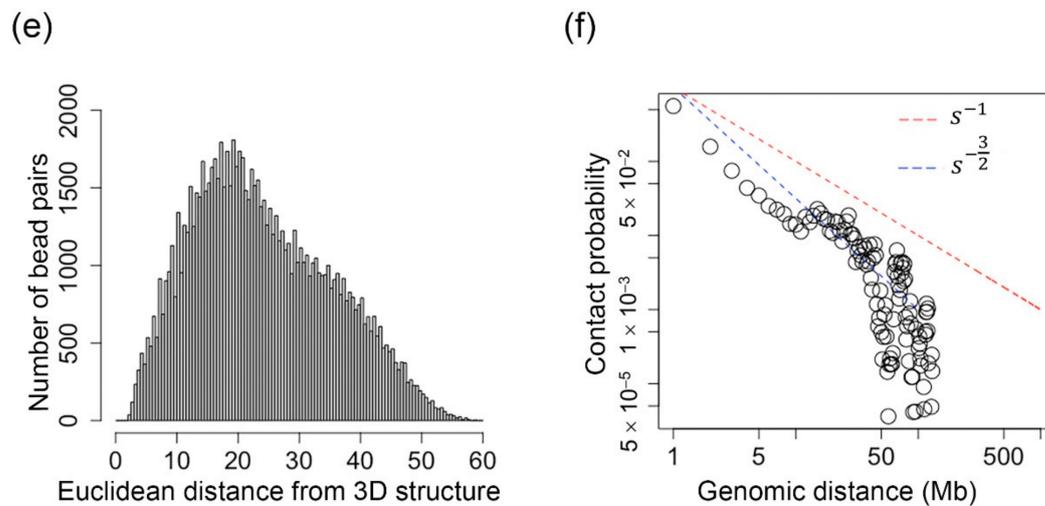


Figure 1. (a) The inferred 3D structure of the X-chromosome of a mouse Th1 cell at 50 kb resolution when n in Equation (3) equals 2. (b) Each black dot indicates one single-cell Hi-C contact. The darker orange color indicates higher Euclidean distances parsed from the inferred 3D structure, and the white color indicates small Euclidean distances parsed from the inferred 3D structure. (c) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} = 1$. (d) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} \in (0.7, 1)$. (e) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} \in (0, 0.7]$. (f) The contact probability of bead pairs over different values of genomic distances.

Figure 1c–e show the number of bead pairs that have different ranges of Euclidean distances based on the reconstructed 3D structures when the $\theta_{i,j}$ value is 1, (0.7, 1) and (0, 0.7], respectively. We performed a Mann–Whitney U test, and the p-value for the distributions in Figure 1c,d is 4.1×10^{-24} , the p-value for the distributions in Figure 1c,e is 5.2×10^{-298} , and the p-value for the distributions in Figure 1d,e is 7.2×10^{-69} . These small p-values indicate that the distributions are statistically different. We can see that for all the structures shown in Figure 1, the bead pairs that have $\theta_{i,j} = 1$ usually have much shorter Euclidean distances compared to the bead pairs that have $\theta_{i,j}$ values in the ranges of (0.7, 1) and (0, 0.7]. The value of $\theta_{i,j}$ is directly related to single-cell Hi-C contacts, that is, when there is a single-cell Hi-C contact existing for a bead pair, its $\theta_{i,j} = 1$, and a small $\theta_{i,j}$ value indicates that the bead pair is far away from the other bead pairs that do have single-cell Hi-C contacts. Therefore, the findings in Figure 1c–e indicate that our reconstructed 3D structures well fit the single-cell Hi-C contact data.

Figure 1f shows the relationship between genomic distance, which is represented as s , and the contact probability on a logarithmic scale. For comparison purposes, we plotted two straight lines indicating s^{-1} and $s^{-\frac{3}{2}}$, which are the hallmarks of a fractal globule structure and an ideal equilibrium globule structure, respectively [75]. The fractal globule packing of chromosomes was proposed in [90] and [91] based on population-based Hi-C data. Figure 1f shows that the reconstructed 3D structure is more similar to an equilibrium globule structure with genomic distance <10 Mb and in between fractal and equilibrium globule structures when genomic distance is larger than 20 Mb and smaller than 40 Mb. Our results indicate that the single-cell chromosomal structures have a more complicated or mixed packing of both fractal and equilibrium globules, which also fits the finding from another research [75].

To test the influence of the well depth in the Lennard-Jones potential on the reconstructed 3D chromosomal structures, we benchmarked four different values of n in Equation (3). In our newly-designed loss function, the depth of the energy well of the Lennard-Jones potential is related to $\theta_{i,j}^n$, and the value of $\theta_{i,j}$ is between zero and one, i.e., $0 < \theta_{i,j} \leq 1$, indicating either the existence of a single-cell Hi-C contact between

bead i and j or an imputed value based on a 2D Gaussian function. Details can be found in Materials and Methods. Therefore, the value of n , which is an integer, will generate different outcomes when $\theta_{i,j} = 1$ or $\theta_{i,j} < 1$. For example, when $\theta_{i,j} = 1$, no matter what value we assign to n , $\theta_{i,j}^n$ will always be 1. However, when $\theta_{i,j} = 0.1$, a larger n will make $\theta_{i,j}^n$ much smaller.

Figure 2 shows the inferred 3D structures for the TH1 X-chromosome with $n = 1, 2, 3$, and 4. We have also tested two different values of r_m in Equation (3). Particularly, Figure 2a–d shows four inferred structures with $n = 1, 2, 3$, and 4 and $r_m = 8.976$, and Figure 2e–h are four inferred structures with $n = 1, 2, 3$, and 4 and $r_m = 8$. It can be found that the overall topologies of the inferred structures are stable, particularly when $n = 1, 2$, and 3. We calculated the Pearson's correlation between $\theta_{i,j}$ values and pairwise Euclidean distances parsed from these inferred 3D structures, which will be discussed in detail in the next section.

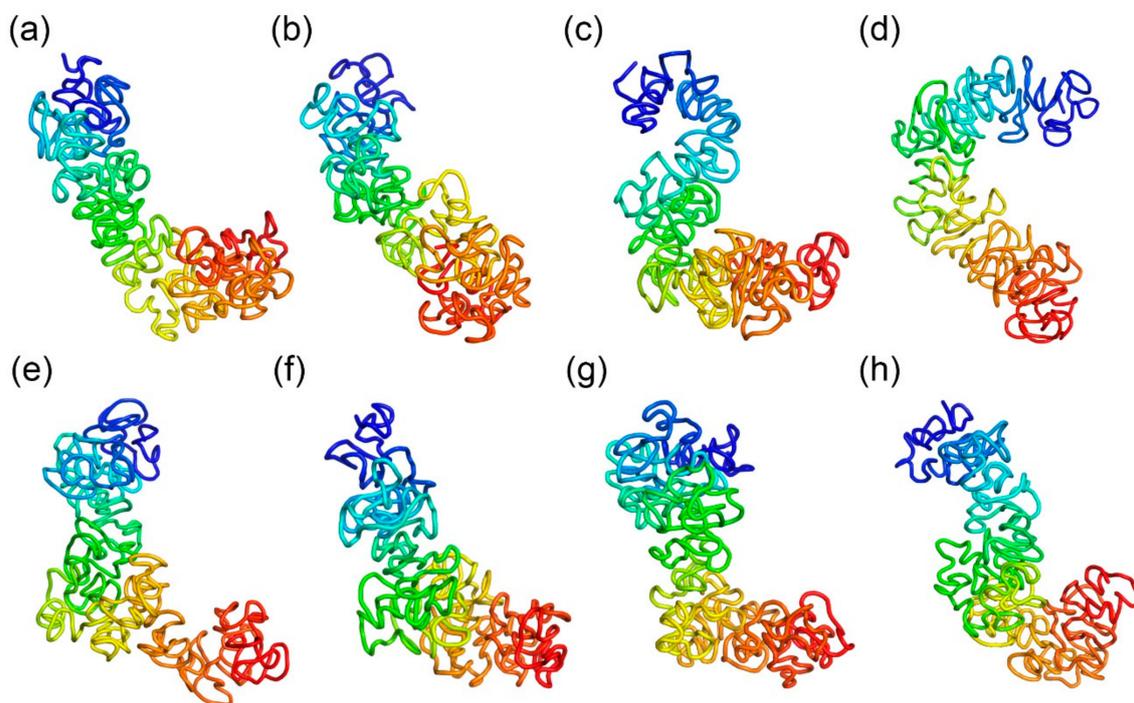


Figure 2. (a–d) are the 3D structures of the X-chromosome of a TH1 cell when r_m in Equation (3) equals 8.976 and n in Equation (3) equals 1, 2, 3, and 4, respectively. (e–h) are the 3D structures of the X-chromosome of a TH1 cell when r_m in Equation (3) equals 8 and n in Equation (3) equals 1, 2, 3, and 4, respectively.

2.2. Pearson's Correlation Between $\theta_{i,j}$ and Euclidean Distances Parsed from the Inferred 3D Structures

We tested different values of n in Equation (3) including 1, 2, 3, and 4 on the same mouse TH1 cell, and then calculated the Pearson's correlation coefficients between the $\theta_{i,j}$ values and the Euclidean distances parsed from the inferred 3D structures. A higher $\theta_{i,j}$ value, that is, closer to 1, indicating a single-cell Hi-C contact between the bead pairs or being adjacent to other bead pairs that have single-cell Hi-C contacts, should correspond to a smaller Euclidean distance between the bead pairs in the reconstructed 3D structure. Therefore, a negative correlation is expected, and the smaller the correlation value is, the better the reconstructed 3D structures fit the $\theta_{i,j}$ values.

The Pearson's correlations for the 3D structures shown in Figure 2a–d are -0.15 , -0.21 , -0.15 , and -0.02 , respectively. The p -values of these correlation values are 0.0 except for the last one (-0.02) being 10^{-7} . The Pearson's correlations for the 3D structures shown in Figure 2e–h are -0.02 , -0.26 , -0.14 , and -0.16 , respectively. The p -values of

these correlation values are 0.0 except for the first one (-0.02) being 10^{-7} . These Pearson's correlation values were calculated on the structures at 500 kb resolution, and each of the 3D structures consists of 333 beads. Therefore, the Pearson's correlations were calculated on 55278 bead pairs since we used half of the $\theta_{i,j}$ and distance matrices and excluded the beads on the diagonal lines. This makes all of the correlation values statistically significant, although the absolute values of the correlations are not very high. Since the largest correlation value is achieved when $n = 2$, we used $n = 2$ for all of the other structures, and 2 is used as the default value for n in our software.

2.3. Comparison with Existing Tools

We compared the 3D structures inferred by our method with the structures inferred by SCL [77] and nuc_dynamics [74]. Figure 3a shows the 3D structure inferred from the method presented in this paper. Figure 3d,g are the 3D structures of the same chromosome but inferred by SCL and nuc_dynamics, respectively. All of these structures were inferred based on the same single-cell Hi-C data of the X-chromosome of a mouse TH1 cell as used in Figures 1 and 2. For Figure 3c,f, two beads being in contact is defined as having a Euclidean distance < 8 , and for Figure 3i, it is a Euclidean distance < 0.5 since the structures inferred by different tools are in different scales.

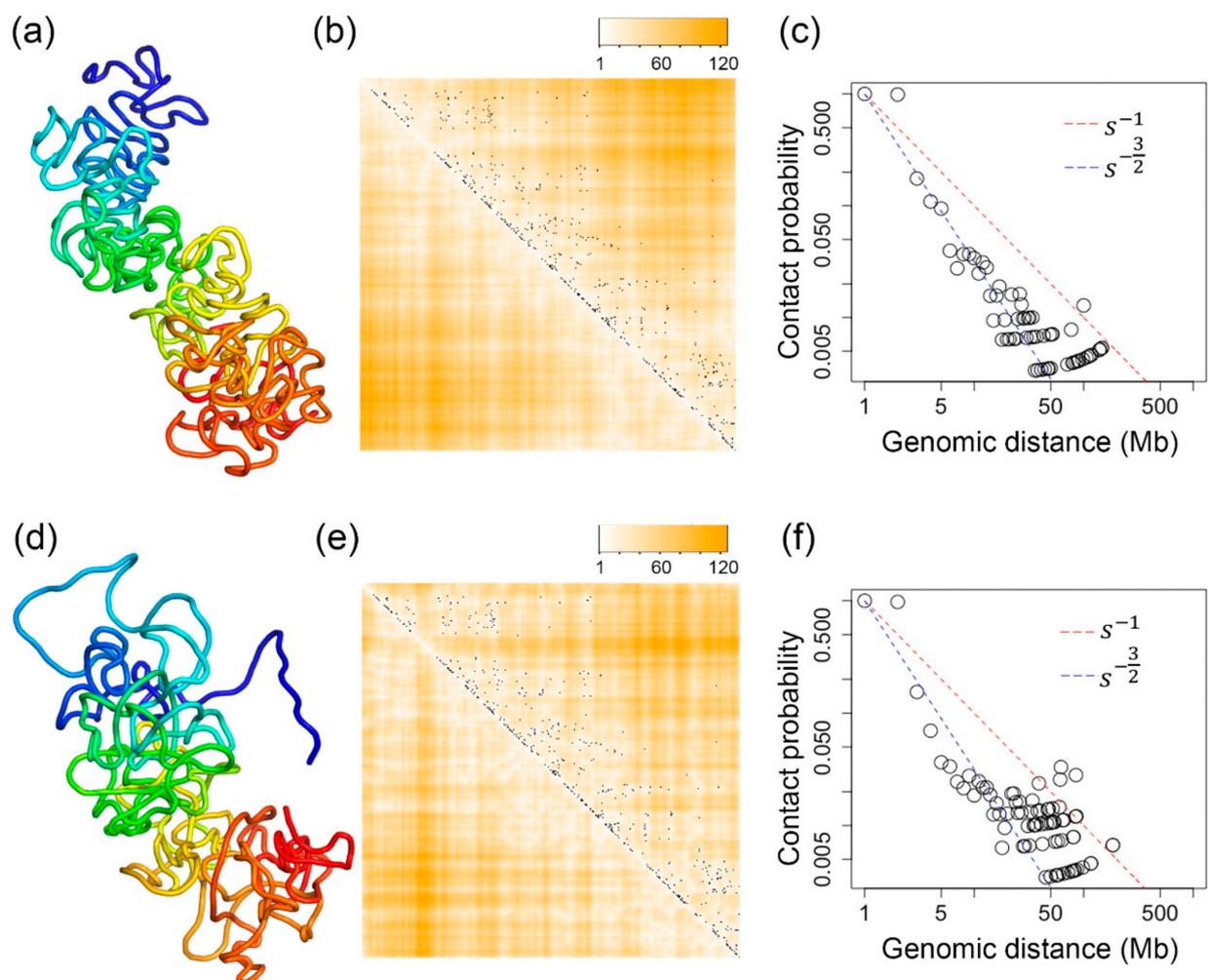


Figure 3. Cont.

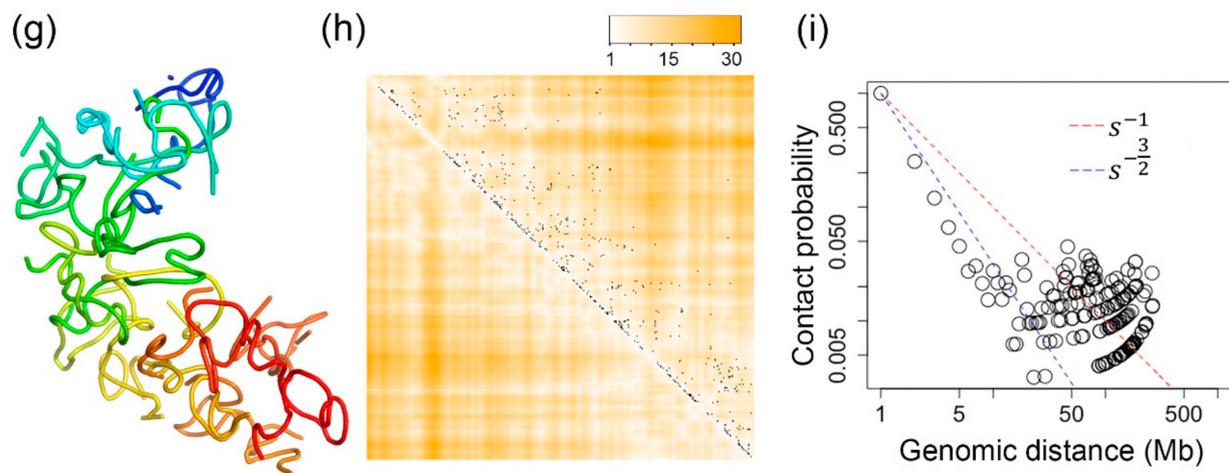


Figure 3. (a–c) The 3D structure of the X-chromosome of a TH1 cell that is inferred by our Lennard-Jones method and its evaluations. (d–f) The 3D structure of the X-chromosome of a TH1 cell that is inferred by SCL and its evaluations. (g–i) The 3D structure of the X-chromosome of a TH1 cell inferred by nuc_dynamics and its evaluations. (a,d,g) are the 3D structures inferred by our Lennard-Jones method, SCL, and nuc_dynamics. (b,e,h) Each black dot indicates one single-cell Hi-C contact. The darker orange color indicates higher Euclidean distances parsed from the inferred 3D structure, and the white color indicates small Euclidean distances parsed from the inferred 3D structure. (c,f,i) The contact probability of bead pairs over different values of genomic distances.

Figure 4 shows the number of bead pairs that have different ranges of Euclidean distances based on the reconstructed 3D structures when $\theta_{i,j}$ is 1, (0.7, 1) and (0, 0.7). We performed a Mann–Whitney U test, and the p -value for the distributions in Figure 4a,b is 9.5×10^{-22} , the p -value for the distributions in Figure 4a,c is 0, and the p -value for the distributions in Figure 4b,c is 3.4×10^{-122} . The p -value for the distributions in Figure 4d,e is 1.24×10^{-28} , the p -value for the distributions in Figure 4d,f is 0, and the p -value for the distributions in Figure 4e,f is 0. The p -value for the distributions in Figure 4g,h is 9.8×10^{-18} , the p -value for the distributions in Figure 4g,i is 1.49×10^{-266} , and the p -value for the distributions in Figure 4h,i is 5.85×10^{-75} . These small p -values indicate that these tools can model the bead pairs with different range of $\theta_{i,j}$ in a statistically different way.

Two large regions in the X-chromosome have no single-cell Hi-C data. One of these two regions exists at the beginning of the X-chromosome, which is corresponding to beads 1–11 (the dark blue region in the inferred 500 kb resolution 3D structures), whereas the other region is corresponding to beads 48–66 (the light blue region in the inferred 3D structures). Single-cell Hi-C experiments cannot work at these telomere or centromere regions. Therefore, the single-cell Hi-C data related to these regions are missing. Although SCL also uses the same 2D Gaussian function to impute for missing Hi-C values, the 3D structures for these two regions inferred by SCL are not folded—see Figure 3d. For nuc_dynamics, the software directly removes the 3D structures for this type of region. This is why the 3D structure for the X-chromosome is not continuous in Figure 3g. However, our Lennard-Jones-based method can smoothly fold the structures for these regions—see Figure 3a.

2.4. Inferred 3D Structures of the Active and Inactive X-Chromosomes of a Human GM12878 Cell

We reconstructed and evaluated the 3D structures of the two X-chromosomes of a human GM12878 cell. The single-cell Hi-C data were downloaded from [73] with file name GSM3271349_gm12878_03.clean.con.txt.gz. The paper also released imputed single-cell Hi-C data. However, to test the ability of the tools to handle extremely sparse single-cell Hi-C data, we used the original Hi-C data—see Figure 5c for example.

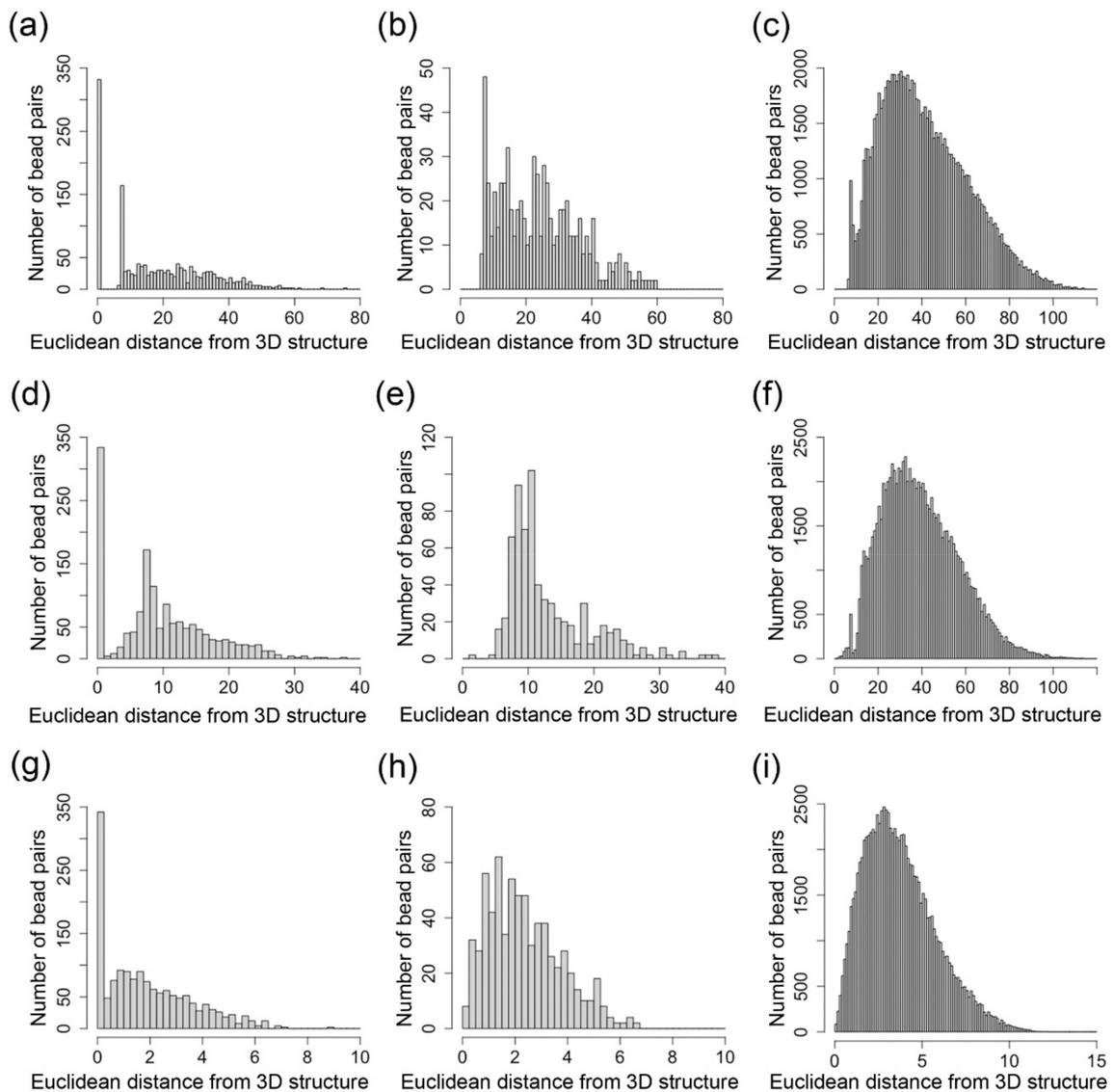


Figure 4. (a–c) Distributions of the number of nodes for the 3D structure inferred by our Lennard-Jones method. (d–f) Distributions of the number of nodes for the 3D structure inferred by SCL. (g–i) Distributions of the number of nodes for the 3D structure inferred by nuc_dynamics. (a,d,g) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} = 1$. (b,e,h) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} \in (0.7, 1)$. (c,f,i) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} \in (0, 0.7]$.

The structures in Figure 5a,e are for the inactive X-chromosome, inferred by our Lennard-Jones method and SCL, respectively. The structures in Figure 5i,m are for the active X-chromosome, inferred by our Lennard-Jones method and SCL, respectively. Figure 5b,f,j,n show the 50 kb resolution structures corresponding to Figure 5a,e,i,m. The inactive X-chromosome is usually silenced in terms of gene expression. The transcripts of lncRNA Xist spread onto one of the two X-chromosomes of female human cells and then alter the 3D structure of the X-chromosome into a highly compacted structure so that the transcriptions on that X-chromosome are stopped [92]. To verify whether the reconstructed 3D structures fit the biological meaning, we calculated the radius of gyration of the 3D structures of inactive and active X-chromosomes. The radius of gyrations of inactive and active structures inferred by our Lennard-Jones method, i.e., Figure 5a,i, are 0.42 and 0.45, respectively. The radius of gyration is a measure of the compactness of a 3D structure. The

smaller radius of gyration for the 3D structure of inactive X-chromosome indicates higher compactness of that structure, which fits the biological meaning of inactive X-chromosome and further proves the correctness of our reconstructed 3D structures.

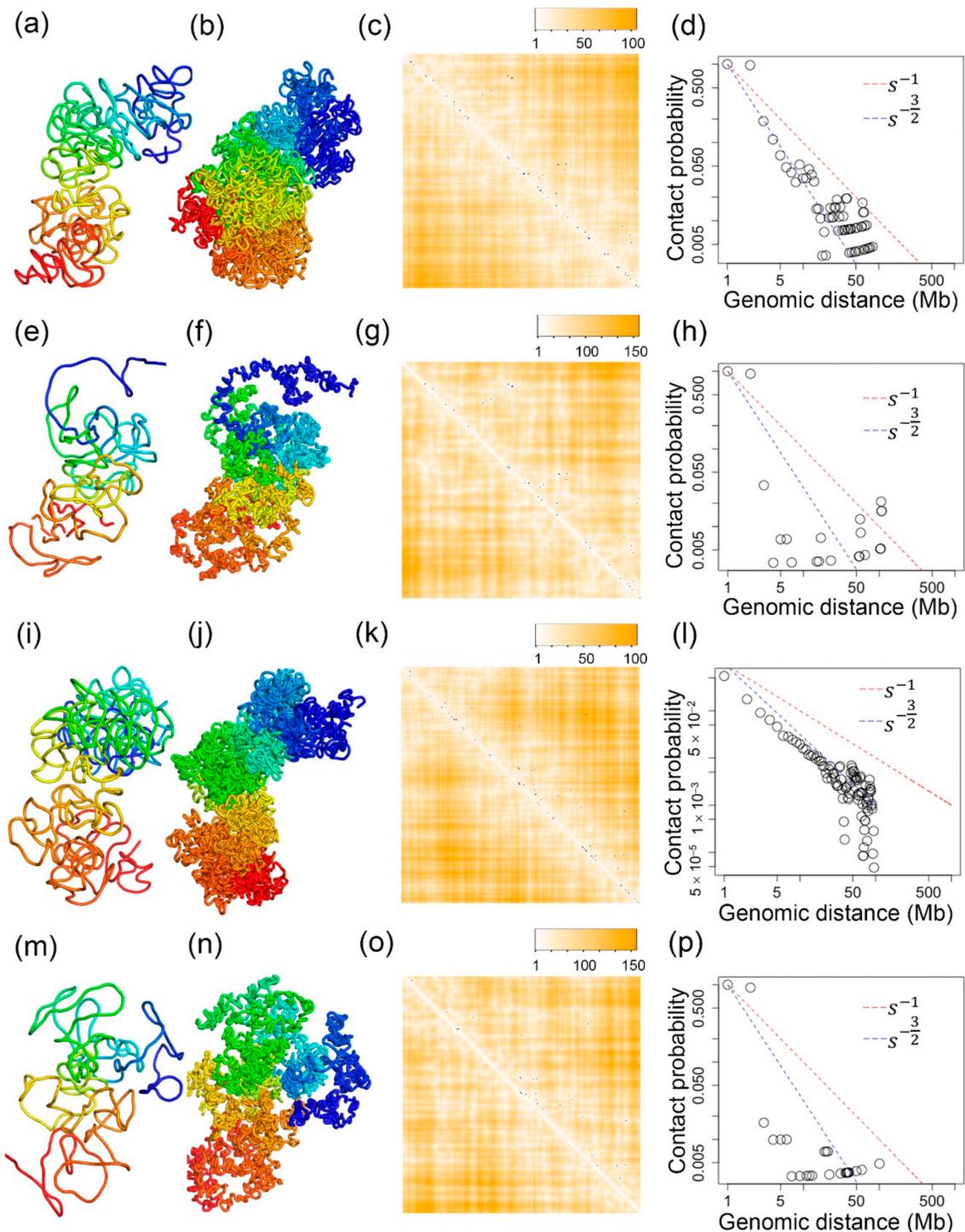


Figure 5. (a,i) The 500 kb resolution 3D structures of the inactive and active X-chromosome of a GM12878 cell, which is inferred by our Lennard-Jones method, respectively. (e,m) The 500 kb resolution 3D structures of the inactive and active X-chromosome of a GM12878 cell, which is inferred by SCL, respectively. (b,f,j,n) are the 50 kb resolution structures corresponding to (a,e,i,m). (c,d) are the evaluations for (a). (g,h) are the evaluations for (e). (k,l) are the evaluations for (i). (o) and (p) are the evaluations for (m).

From Figure 5d,h,l,p, it can be found that for extremely sparse single-cell Hi-C data, the 3D structures generated by our Lennard-Jones method still maintain fractal or equilibrium globule structures—see Figure 5d,l—whereas the structures generated by SCL have deviated from fractal or equilibrium globule structures—see Figure 5h,p. This indicates the advantage of our Lennard-Jones method for generating stable structures based on extremely sparse single-cell Hi-C data.

2.5. Inferred 3D Structures of the Chromosome 3 of a Mouse Oocyte Cell

Figure 6 shows the inferred 3D structure and its evaluations for the chromosome 3 of a mouse oocyte cell. The single-cell Hi-C data was downloaded from [52] (144_oocyte_SN). We performed a Mann–Whitney U test for the distributions of the number of bead pairs with different $\theta_{i,j}$ values, and the p-value for the distributions in Figure 6c,d is 2.6×10^{-27} , the p-value for the distributions in Figure 6c,e is 0, and the p-value for the distributions in Figure 6d,e is 0.

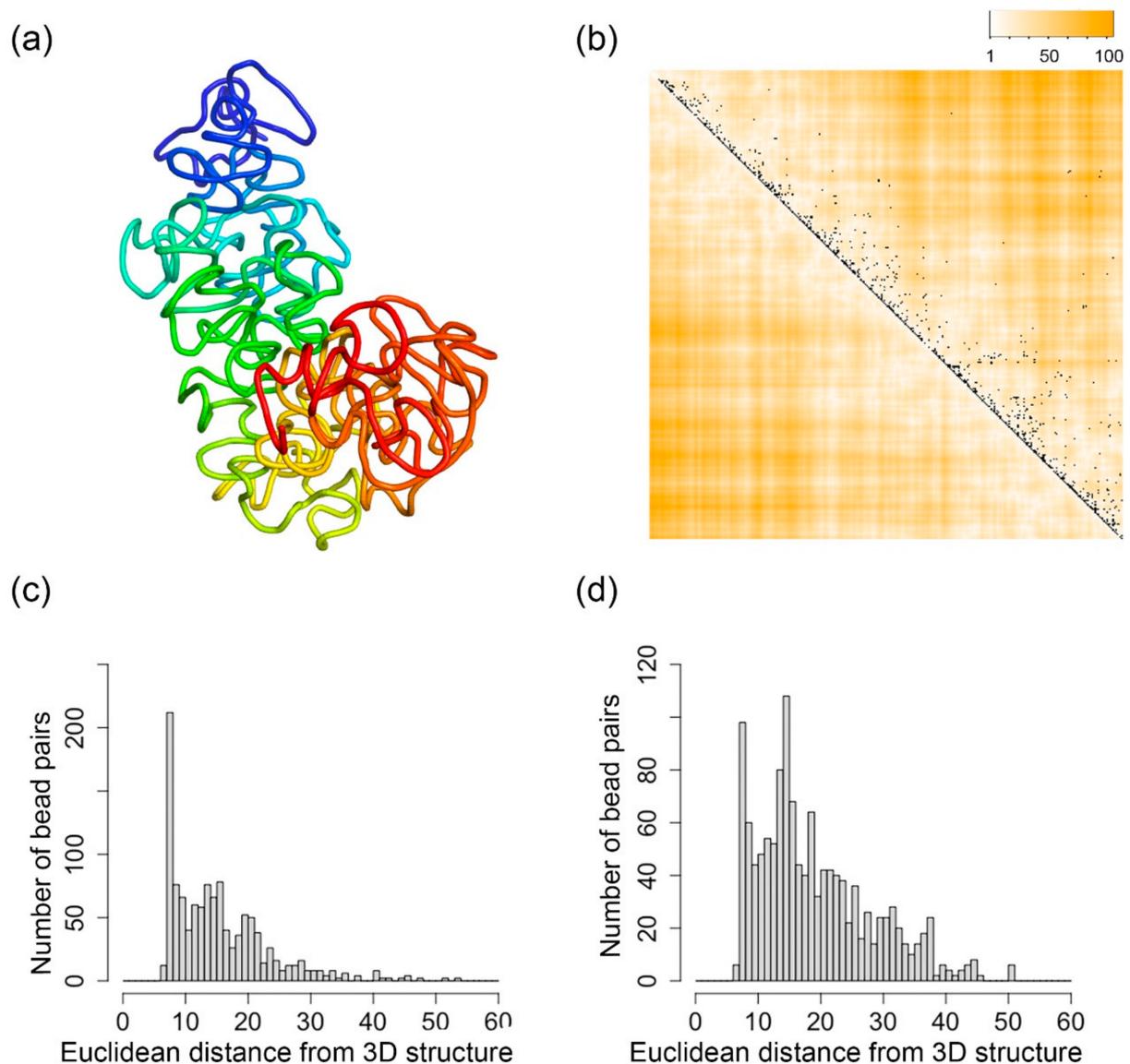


Figure 6. Cont.

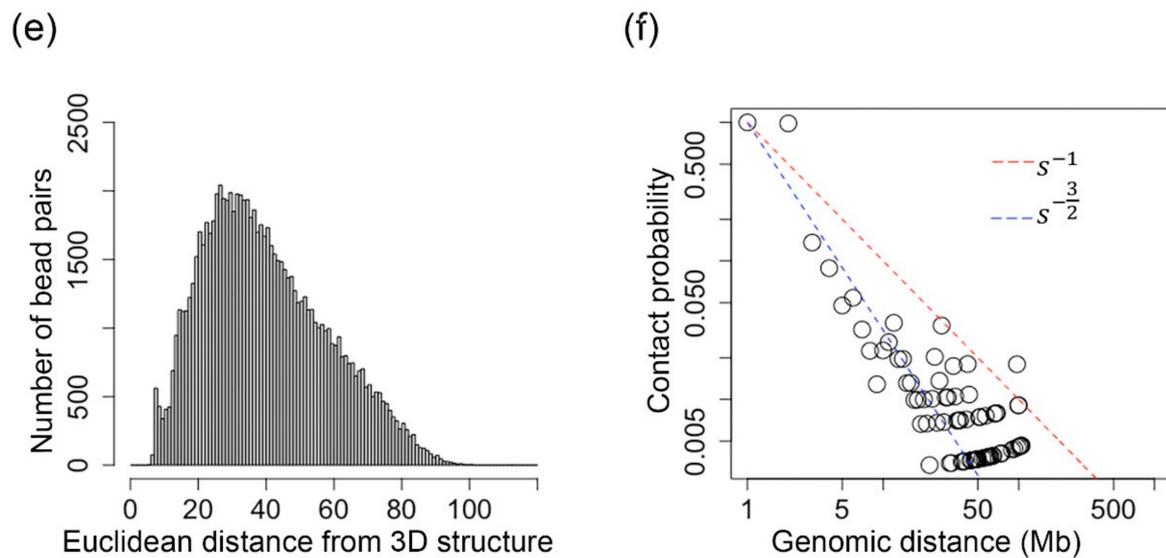


Figure 6. (a) The inferred 3D structure of the chromosome 3 of a mouse oocyte cell at 500 kb resolution. (b) Each black dot indicates one single-cell Hi-C contact. The darker orange color indicates higher Euclidean distances parsed from the inferred 3D structure, and the white color indicates small Euclidean distances parsed from the inferred 3D structure. (c) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} = 1$. (d) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} \in (0.7, 1)$. (e) The distribution of the number of bead pairs over different values of Euclidean distances parsed from the inferred structure when $\theta_{i,j} \in (0, 0.7]$. (f) The contact probability of bead pairs over different values of genomic distances.

2.6. Validation with 3D-FISH

3D Fluorescence in situ hybridization (FISH) is a type of biochemistry experiment that can detect the distance between pairs of genomic regions. Therefore, by comparing our inferred 3D chromosomal structures with the results of 3D-FISH, we can evaluate the accuracy of the inferred 3D genome structures. We calculated the Pearson's correlation between the distances of eight probe pairs parsed from the inferred 3D structures and the distances between the same pairs of eight probes detected by 3D-FISH [93]. These eight probe pairs are located on chromosomes 3 and 11. The evaluation was conducted on the reconstructed 3D structure of the TH1 cell at 500 kb resolution. Higher Pearson's correlation values here indicate that the inferred 3D structure better fits the results from 3D-FISH. We found that when $n = 1$ and $r_m = 8.976$ the 3D structure inferred by our Lennard-Jones method has a Pearson's correlation of 0.22 to the 3D FISH data.

3. Discussion

This research explores the possibility of using the Lennard-Jones potential to reconstruct the 3D structures of chromosomes based on single-cell Hi-C data. Compared to population-cell Hi-C data, single-cell Hi-C contact data are very sparse. This makes it a challenging computational problem because traditional algorithms usually need to convert the number of Hi-C contacts into a target distance, and zero contact would be converted to infinity.

Some studies use Lennard-Jones to model the 3D structures of proteins, e.g., [89]. In those studies, the Lennard-Jones potential was used to model the attractive and repulsive forces between different types of protein residues. However, when we model the 3D structure of chromosomes, each bead is considered to represent m DNA base pairs, where m is the resolution that can be 500 kb or 50 kb in this research. These DNA bead pairs may not have the clearly defined attractive or repulsive forces as residues in a protein have. Therefore, it may not be possible to infer chromosomal 3D structures by directly modeling the attractive or repulsive forces between bead pairs using the Lennard-Jones potential.

Therefore, we used the well depth of the Lennard-Jones potential to model the confidence for a bead pair to maintain an in-contact distance or the stability of the in-contact relationship. For the bead pairs that we observe a single-cell Hi-C contact for, our loss function assigns stronger confidences or deeper well depths in the Lennard-Jones potential. This makes these bead pairs have a strong potential to maintain such a short distance during the Monte Carlo simulations. For the bead pairs that are far away from the bead pairs that do have single-cell Hi-C contacts, our loss function assigns lower confidences by assigning shallower well depths in the Lennard-Jones potential. The degree of being smaller can be controlled by the value of n in Equation (3): a larger n will make the well depth increasingly smaller. In other words, for the bead pairs that have small $\theta_{i,j}$ values, we do not force a long distance between the beads, but make their in-contact distances harder to be maintained or easier to be broken in the Monte Carlo simulations.

The results shown in this study demonstrated that this strategy worked. We tested different values of n influencing the well depth in the Lennard-Jones potential and conducted evaluations from multiple perspectives. We also cross-validated the inferred 3D structures with the data obtained by 3D-FISH, another biochemistry experiment that can detect 3D proximities of multiple DNA segments. We applied our methods to active and inactive X-chromosomes in a female human cell. The reconstructed 3D structure for the inactivate X-chromosome has a smaller radius of gyration, which fits the biological meaning of X-chromosome inactivation. We also found that our method can keep the inferred 3D structure a mix of fractal and equilibrium globule structure even for extremely sparse single-cell Hi-C data.

4. Materials and Methods

4.1. Introduction to Negative Potential Energy

The negative potential energy is explained by the three red balls in three different gravitational potential energy statuses in Figure 7. In a flat horizontal plane, the red ball will be stable, that is, it will not move unless pushed by a source of external force that provides some kinetic energy. This situation is called zero energy status. If a ball is located on the top of the hill, gravity will make the ball start rolling down at an increasing speed. During this process, the red ball would emit potential energy so that this situation is labeled as a positive energy status. In the well, the red ball needs to absorb some kinetic energy to start moving and eventually jumps out of the well when it gets enough kinetic energy.

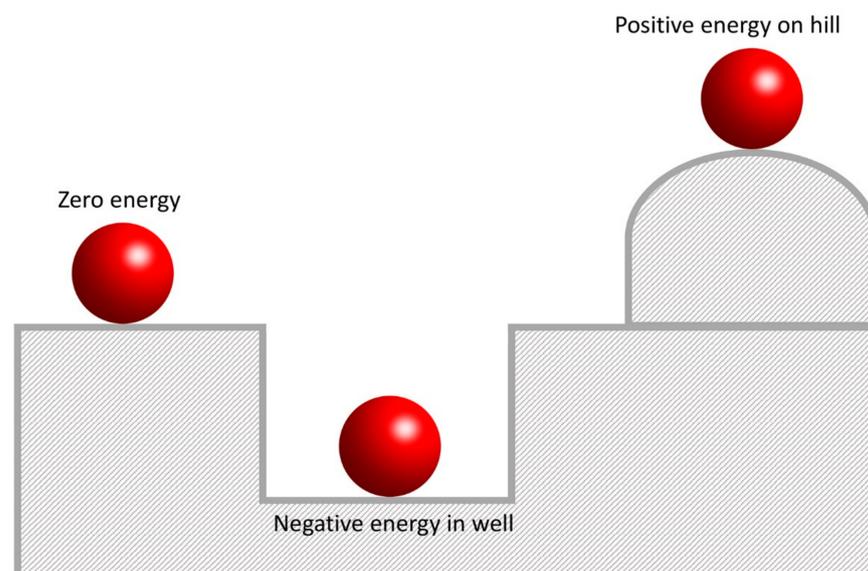


Figure 7. A diagram showing the negative, positive, and zero energy.

In Figure 8a, the two balls do not interact when there is a long distance between them. In Figure 8b, when the two balls are brought closer to a distance r , they will start to attract each other, i.e., there is an attractive force between them. In Figure 8c, at an even closer distance, the two balls will first accelerate towards each other due to the dominant attractive force, and then reach an equilibrium distance at which the minimum bonding potential is reached. In Figure 8d, when the two balls are getting further closer to each other (for example, pushed by an external source of energy), i.e., the distance between them is less than the equilibrium distance described in (c), the repulsive force will dominate and keeps increasing to try to push the two balls away.

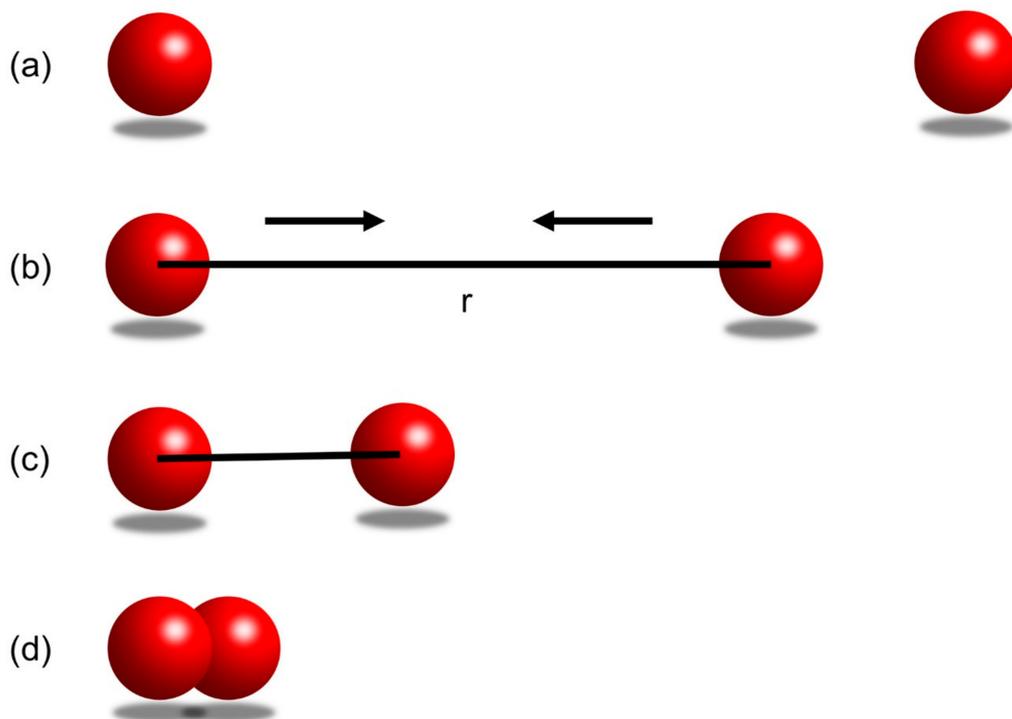


Figure 8. (a) The balls are not interacting as they have an infinite distance. (b) Two balls are brought closer with minimum energy to a distance of r . The two balls have an attractive force between them. (c) The two balls are brought closer by the attractive force between them until they reach an equilibrium distance, at which they reach the minimum bonding potential. (d) External energy pushes the two balls even closer. A repulsive force tries to push the two balls apart, and the repulsive force is greater than the attractive force.

4.2. The Lennard-Jones Potential

The Lennard-Jones potential is a mathematical formula that can be used to model the potential energy and interactions between two atoms or molecules. Figure 9 shows the curve of the Lennard-Jones potential and the repulsive and attractive forces. The x -axis in the plot indicates the distance between two particles, and the y -axis indicates the inter-particle potential or potential energy. As shown in Figure 9, when the distance between two particles is at the equilibrium distance or r_m , the force is zero, and the potential energy reaches its minimum, that is, it is at the bottom of the energy well. When the distance increases from the equilibrium distance, the force between the two particles will be the attractive force. Meanwhile, the potential energy increases and gets closer to zero. When the distance decreases from the equilibrium distance, the repulsive force will be gradually stronger.

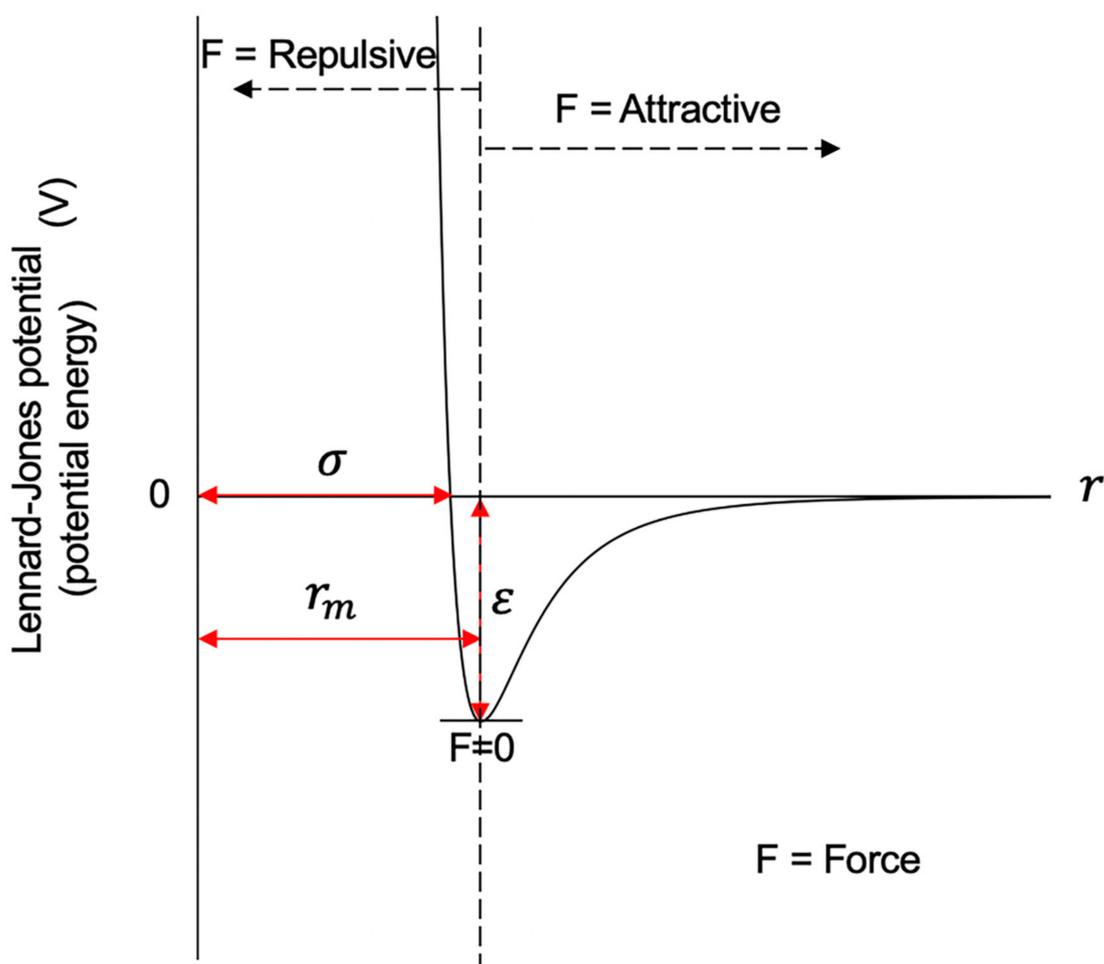


Figure 9. A diagram showing the curve of the Lennard-Jones potential.

The depth of the well indicates how strongly the two particles attract each other. As shown in Figure 9, the deeper the well is, the more external energy the ball would need to jump out of the well, that is, the harder it is to get the ball out of the well. Similarly, the deeper the well of the Lennard-Jones potential is, the more stable the conformation between the two balls is. This is the key feature that we have used in inferring 3D chromosomal structures, which will be discussed later in detail.

The mathematical definition of the Lennard-Jones potential is:

$$V_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] = \epsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right] \quad (1)$$

in which ϵ is the depth of the potential well, σ represents the distance at which the potential equals zero, r is the distance between the two beads, and r_m is the distance at which the energy reaches the bottom of the energy well and $r_m = 2^{\frac{1}{6}}\sigma \approx 1.122\sigma$. In other words, the value of the potential function is $-\epsilon$ when the distance is r_m .

4.3. 2D Gaussian Function for Imputing Single-Cell Hi-C Contacts

The single-cell Hi-C contact matrix is sparse, which means that the number of 0s is much bigger than the number of 1s in the contact matrix. For example, there are ~400 1s in the single-cell Hi-C contact matrix for chromosome 1 of the mouse TH1 dataset while all other cells in the matrix are 0s. Each of the 1s indicates that there is at least one single-cell Hi-C contact observed between the bead pairs.

We used a 2D Gaussian function to impute the dominating zero values in the contact matrix as we previously did in [77], i.e., transform the 0s into non-zero values based on the distances to the location where a single-cell Hi-C contact is observed. The intuition behind this is that if beads i and j are in contact in the 3D space, the sequentially adjacent beads of these two beads should not be far away. For example, if beads i and j are proximate in the 3D space, the sequentially adjacent beads such as beads $i - 1$ and $i + 1$ should also be spatially close to, or at least not far away from, beads $j, j + 1$, and $j - 1$. The 2D Gaussian function we have used in this research is:

$$\theta_{i,j} = \sum_{\substack{Hi-C(x_p, y_p) = 1 \\ |x_p - i| \leq d_0 \text{ and } |y_p - j| \leq d_0}} \exp\left(-\left(\frac{(x_p - i)^2}{\mu_2} + \frac{(y_p - j)^2}{\mu_2}\right)\right) \quad (2)$$

in which x_p, y_p indicate the location of a single-cell Hi-C contact in the contact matrix and parameter μ_2 can be used to control the shape of the 2D Gaussian function. The larger the parameter μ_2 is, the flatter the Gaussian-like curve would be, which means that the more bead pairs surrounding the single-cell Hi-C contact can be influenced. The $(x_p - i)$ and $(y_p - j)$ indicate the x -axis and y -axis distances from a zero value to the location that has a single-cell Hi-C contact. We use a parameter d_0 to control how far the influence of a single-cell Hi-C contact can reach.

4.4. Loss Function

The following Equation (3) is the loss function that we newly designed based on the Lennard-Jones potential in Equation (1):

$$E_{i,j} = \beta \times \theta_{i,j}^n \left[\left(\frac{r_m}{d_{i,j}}\right)^{12} - 2 \left(\frac{r_m}{d_{i,j}}\right)^6 \right] \quad (3)$$

in which β is a scaling factor and set to 10, $\theta_{i,j}$ is the output of the 2D Gaussian function, $d_{i,j}$ is the distance between each bead pair in the 3D structure, and the value $r_m = 2^{\frac{1}{6}}\sigma \approx 1.122\sigma$. We tested two different values for r_m . The first way is to set $r_m = 8$, which means if two beads have their $\theta_{i,j} = 1$, our algorithm will try to make their distance be 8 in the inferred 3D structure. Notice that SCL also uses 8 as the default value for the bead pairs having $\theta_{i,j} = 1$. In this way, we can objectively compare the performance of our method and SCL. The second way is to set $\sigma = 8$, which means $r_m = 1.122 \times 8 = 8.976$. If $d_{i,j}$ or the distance between any bead pairs is >16 , we will not calculate the loss value using Equation (3) but directly assign a zero to the bead pair, that is, $E_{i,j} = 0$. This is a similar idea to truncated Lennard-Jones potential, which can save computational time.

4.5. Initialization of Random 3D Chromosomal Structures

We used the same initialization step as in SCL [77]. Each chromosome is represented by a beads-on-a-string approach, that is, a chromosome is evenly divided into beads, and each DNA bead consists of a fixed number of DNA nucleotides or base pairs, e.g., 500 kb. A 3D lattice was created with the size of each side equaling $5l$ with l being the number of DNA beads for the whole chromosome. For example, the X-chromosome of the mouse contains 333 beads, making the size of the 3D lattice $(5 \times 333) \times (5 \times 333) \times (5 \times 333)$. Each small cubic in the 3D lattice has the size of $1 \times 1 \times 1$. The 3D structure of a chromosome was initialized by randomly adding DNA beads one by one in a way that the newly added DNA bead must have a distance of $[2,10]$ with $\sqrt{8}$ excluded [94] with all other currently existing beads in the 3D cubic lattice.

4.6. The Metropolis–Hastings Simulations

The cooling schedule defined in [95] was used in the simulated annealing process. The initial temperature was set as $T_0 = 10$. The current temperature during each iteration of the simulated annealing process is $T_c = T_0 \times (0.9)^p$ with p representing the number of times that the temperature has been decreased.

To stabilize the system at a certain temperature, a sufficient number of attempts were allowed at each temperature. If the number of attempts at a certain temperature reaches 100 times the number of beads or there are 10 accepted moves for each DNA bead on average [95], the temperature decreases based on the cooling schedule, followed by a new iteration of attempts. During each attempt, the Metropolis–Hastings algorithm was used to randomly select a DNA bead to randomly move from its current location to one of its 26 adjacent cells in the 3D cubic lattice (nine from the level above the current location, nine from the level below the current location, and eight from the same level of current location). The acceptance of move depends on ΔLoss , which is the change of the value of the loss function before and after the move: $\Delta\text{Loss} = \text{Loss}(\text{after}) - \text{Loss}(\text{before})$. If $\Delta\text{Loss} \leq 0$, the move is accepted all the time. If $\Delta\text{Loss} > 0$, the move is accepted with a probability $P = e^{-\Delta\text{Loss}/T_c}$, where T_c indicates the current temperature [95].

To generate high-resolution structures with 50 kb resolution, the 3D structure with 500 kb resolution was generated first. After that, nine beads were added in between every two consecutive 500 kb beads. These nine beads were added in the same way as in the initialization step. After that, we only applied 51 attempts with a constant temperature of 0.1, which is only to refine the structure without a large alternation.

4.7. Model Selection

A total of 50 structures were generated, and the one with the highest quality was selected and shown in the figures. We used TM-score to compare each of the models against all of the others and then calculated the average structural similarity. For each model, an average structural similarity is calculated. All models were ranked based on the average structural similarity, and the model ranked at No.1 or which had the highest average structural similarity was selected and used as the final output.

4.8. Computational Time

We implemented the system in C++, and the source code of the system can be found at <http://dna.cs.miami.edu/LJ3D/> (accessed on 30 May 2021). Benchmarked on an Intel Xeon CPU at 2.70 GHz, it took our Lennard-Jones method 7 min to model the 3D structure of the X-chromosome of a TH1 cell at 500 kb resolution, and 2 h 25 min for the structure at 50 kb resolution. In comparison, SCL needs 27 min and 3 h 17 min to model the structure of the same chromosome at 500 kb and 50 kb resolutions, respectively.

5. Conclusions

The Lennard-Jones potential can be used to infer 3D chromosomal structures based on single-cell Hi-C data. The well depth of the Lennard-Jones potential can be used to model how easy it is to break the distances between bead pairs. For the bead pairs that have a single-cell Hi-C contact observed, our loss function assigns a deeper well that makes it harder to be broken. For the bead pairs that do not have a single-cell Hi-C contact, we used a 2D Gaussian function to impute their contact values based on their distances to the other bead pairs that do have single-cell Hi-C contacts, and then the loss function assigns shallower well depths for those bead pairs. We benchmarked a set of different parameters from various perspectives and found the best parameter that results in the best performance.

Author Contributions: Conceptualization, Z.W. and M.Z.; Methodology, M.Z. and Z.W.; Software, M.Z.; Validation, M.Z.; Formal Analysis, M.Z. and Z.W.; Investigation, M.Z. and Z.W.; Resources, M.Z., Z.W., N.W., and C.Z.; Data Curation, M.Z.; Writing—Original Draft Preparation, M.Z.; Writing—Review and Editing, Z.W., C.Z., and N.W.; Visualization, M.Z.; Supervision, Z.W., N.W., and C.Z.; Project Administration, Z.W., N.W., and C.Z.; Funding Acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [National Institute of General Medical Sciences] grant number [R15GM120650] and [R35GM137974] to ZW. The APC was funded by [R35GM137974].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The software of LJ3D can be found at <http://dna.cs.miami.edu/LJ3D/> (accessed on 30 May 2021).

Acknowledgments: We thank Ras B. Pandey at the University of Southern Mississippi for providing insights on the Lennard-Jones potential and Metropolis–Hastings simulations. We thank Hao Zhu for providing inferred 3D structures of SCL so that we can compare our method with SCL.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bickmore, W.A.; van Steensel, B. Genome architecture: Domain organization of interphase chromosomes. *Cell* **2013**, *152*, 1270–1284. [[CrossRef](#)]
- Sexton, T.; Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **2015**, *160*, 1049–1059. [[CrossRef](#)] [[PubMed](#)]
- Pombo, A.; Dillon, N. Three-dimensional genome architecture: Players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 245–257. [[CrossRef](#)] [[PubMed](#)]
- Therizols, P.; Illingworth, R.S.; Courilleau, C.; Boyle, S.; Wood, A.J.; Bickmore, W.A. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **2014**, *346*, 1238–1242. [[CrossRef](#)] [[PubMed](#)]
- Gonzalez-Sandoval, A.; Towbin, B.D.; Kalck, V.; Cabianca, D.S.; Gaidatzis, D.; Hauer, M.H.; Geng, L.; Wang, L.; Yang, T.; Wang, X. Perinuclear anchoring of H3K9-methylated chromatin stabilizes induced cell fate in *C. elegans* embryos. *Cell* **2015**, *163*, 1333–1347. [[CrossRef](#)] [[PubMed](#)]
- Dixon, J.R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J.S.; Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **2012**, *485*, 376–380. [[CrossRef](#)]
- Liu, T.; Wang, Z. Reconstructing high-resolution chromosome three-dimensional structures by hi-C complex networks. *BMC Bioinform.* **2018**, *19*, 39–50. [[CrossRef](#)]
- Liu, T.; Porter, J.; Zhao, C.; Zhu, H.; Wang, N.; Sun, Z.; Mo, Y.-Y.; Wang, Z. TADKB: Family classification and a knowledge base of topologically associating domains. *BMC Genom.* **2019**, *20*, 1–17. [[CrossRef](#)] [[PubMed](#)]
- De Wit, E.; De Laat, W. A decade of 3C technologies: Insights into nuclear organization. *Genes Dev.* **2012**, *26*, 11–24. [[CrossRef](#)]
- Miele, A.; Dekker, J. Mapping cis- and trans-chromatin interaction networks using chromosome conformation capture (3C). In *The Nucleus*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 105–121.
- Rodley, C.; Bertels, F.; Jones, B.; O’Sullivan, J. Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genet. Biol.* **2009**, *46*, 879–886. [[CrossRef](#)]
- Hagège, H.; Klous, P.; Braem, C.; Splinter, E.; Dekker, J.; Cathala, G.; De Laat, W.; Forné, T. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2007**, *2*, 1722. [[CrossRef](#)] [[PubMed](#)]
- Splinter, E.; Grosveld, F.; de Laat, W. 3C technology: Analyzing the spatial organization of genomic loci in vivo. In *Methods in Enzymology*; Elsevier: Berlin/Heidelberg, Germany, 2003; Volume 375, pp. 493–507.
- De Laat, W.; Dekker, J. 3C-based technologies to study the shape of the genome. *Methods* **2012**, *58*. [[CrossRef](#)]
- Sun, C.; Lu, C. Microfluidics-based chromosome conformation capture (3C) technology for examining chromatin organization with a low quantity of cells. *Anal. Chem.* **2018**, *90*, 3714–3719. [[CrossRef](#)] [[PubMed](#)]
- Hövel, I.; Louwers, M.; Stam, M. 3C technologies in plants. *Methods* **2012**, *58*, 204–211. [[CrossRef](#)] [[PubMed](#)]
- Wei, G.; Zhao, K. 3C-based methods to detect long-range chromatin interactions. *Front. Biol.* **2011**, *6*, 76–81. [[CrossRef](#)]
- Louwers, M.; Splinter, E.; Van Driel, R.; De Laat, W.; Stam, M. Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). *Nat. Protoc.* **2009**, *4*, 1216. [[CrossRef](#)]
- Van de Werken, H.J.; de Vree, P.J.; Splinter, E.; Holwerda, S.J.; Klous, P.; de Wit, E.; de Laat, W. 4C technology: Protocols and data analysis. In *Methods in Enzymology*; Elsevier: Berlin/Heidelberg, Germany, 2012; Volume 513, pp. 89–112.
- Splinter, E.; de Wit, E.; van de Werken, H.J.; Klous, P.; de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. *Methods* **2012**, *58*, 221–230. [[CrossRef](#)]

21. Simonis, M.; Klous, P.; Homminga, I.; Galjaard, R.-J.; Rijkers, E.-J.; Grosveld, F.; Meijerink, J.P.; De Laat, W. High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat. Methods* **2009**, *6*, 837. [[CrossRef](#)]
22. Simonis, M.; Klous, P.; Splinter, E.; Moshkin, Y.; Willemsen, R.; De Wit, E.; Van Steensel, B.; De Laat, W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* **2006**, *38*, 1348–1354. [[CrossRef](#)]
23. De Laat, W.; Grosveld, F. Capture and Characterized co-Localized Chromatin (4C) Technology. 2012. Available online: <https://patents.google.com/patent/US8153373B2/en> (accessed on 30 May 2021).
24. Simonis, M.; Kooren, J.; De Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **2007**, *4*, 895–901. [[CrossRef](#)]
25. Van De Werken, H.J.; Landan, G.; Holwerda, S.J.; Hoichman, M.; Klous, P.; Chachik, R.; Splinter, E.; Valdes-Quezada, C.; Öz, Y.; Bouwman, B.A. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **2012**, *9*, 969. [[CrossRef](#)]
26. Gavrillov, A.; Eivazova, E.; Pirozhkova, I.; Lipinski, M.; Razin, S.; Vassetzky, Y. Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. In *Chromatin Immunoprecipitation Assays*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 171–188.
27. Dostie, J.; Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nat. Protoc.* **2007**, *2*, 988. [[CrossRef](#)]
28. Van Berkum, N.L.; Dekker, J. Determining spatial chromatin organization of large genomic regions using 5C technology. In *Chromatin Immunoprecipitation Assays*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 189–213.
29. Ferraiuolo, M.A.; Sanyal, A.; Naumova, N.; Dekker, J.; Dostie, J. From cells to chromatin: Capturing snapshots of genome organization with 5C technology. *Methods* **2012**, *58*, 255–267. [[CrossRef](#)]
30. Dostie, J.; Richmond, T.A.; Arnaout, R.A.; Selzer, R.R.; Lee, W.L.; Honan, T.A.; Rubio, E.D.; Krumm, A.; Lamb, J.; Nusbaum, C. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **2006**, *16*, 1299–1309. [[CrossRef](#)] [[PubMed](#)]
31. Ferraiuolo, M.A.; Sanyal, A.; Naumova, N.; Dekker, J.; Dostie, J. Mapping chromatin interactions with 5C technology: 5C; a quantitative approach to capturing chromatin conformation over large genomic distances. *Methods* **2012**, *58*. [[CrossRef](#)]
32. Van Berkum, N.L.; Lieberman-Aiden, E.; Williams, L.; Imakaev, M.; Gnirke, A.; Mirny, L.A.; Dekker, J.; Lander, E.S. Hi-C: A method to study the three-dimensional architecture of genomes. *JoVE* **2010**, *39*, e1869. [[CrossRef](#)] [[PubMed](#)]
33. Lévy-Leduc, C.; Delattre, M.; Mary-Huard, T.; Robin, S. Two-dimensional segmentation for analyzing Hi-C data. *Bioinformatics* **2014**, *30*, i386–i392. [[CrossRef](#)] [[PubMed](#)]
34. Yang, T.; Zhang, F.; Yardımcı, G.G.; Song, F.; Hardison, R.C.; Noble, W.S.; Yue, F.; Li, Q. HiCRep: Assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **2017**, *27*, 1939–1949. [[CrossRef](#)]
35. Belton, J.-M.; McCord, R.P.; Gibcus, J.H.; Naumova, N.; Zhan, Y.; Dekker, J. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **2012**, *58*, 268–276. [[CrossRef](#)]
36. Serra, F.; Baù, D.; Goodstadt, M.; Castillo, D.; Fillion, G.; Marti-Renom, M.A. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* **2017**, *13*, e1005665. [[CrossRef](#)] [[PubMed](#)]
37. Zhang, Y.; An, L.; Xu, J.; Zhang, B.; Zheng, W.J.; Hu, M.; Tang, J.; Yue, F. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nat. Commun.* **2018**, *9*, 1–9. [[CrossRef](#)]
38. Peng, C.; Fu, L.-Y.; Dong, P.-F.; Deng, Z.-L.; Li, J.-X.; Wang, X.-T.; Zhang, H.-Y. The sequencing bias relaxed characteristics of Hi-C derived data and implications for chromatin 3D modeling. *Nucleic Acids Res.* **2013**, *41*, e183. [[CrossRef](#)]
39. Stansfield, J.C.; Cresswell, K.G.; Vladimirov, V.I.; Dozmorov, M.G. HiCcompare: An R-package for joint normalization and comparison of Hi-C datasets. *BMC Bioinform.* **2018**, *19*, 279. [[CrossRef](#)] [[PubMed](#)]
40. Hu, M.; Deng, K.; Qin, Z.; Liu, J.S. Understanding spatial organizations of chromosomes via statistical analysis of Hi-C data. *Quant. Biol.* **2013**, *1*, 156–174. [[CrossRef](#)] [[PubMed](#)]
41. Hu, M.; Deng, K.; Selvaraj, S.; Qin, Z.; Ren, B.; Liu, J.S. HiCNorm: Removing biases in Hi-C data via Poisson regression. *Bioinformatics* **2012**, *28*, 3131–3133. [[CrossRef](#)]
42. Lin, D.; Hong, P.; Zhang, S.; Xu, W.; Jamal, M.; Yan, K.; Lei, Y.; Li, L.; Ruan, Y.; Fu, Z.F. Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nat. Genet.* **2018**, *50*, 754–763. [[CrossRef](#)]
43. Wu, H.-J.; Michor, F. A computational strategy to adjust for copy number in tumor Hi-C data. *Bioinformatics* **2016**, *32*, 3695–3701. [[CrossRef](#)]
44. Ning, C.; He, M.; Tang, Q.; Zhu, Q.; Li, M.; Li, D. Advances in mammalian three-dimensional genome by using Hi-C technology approach. *Yi Chuan Hered.* **2019**, *41*, 215–233.
45. Battulin, N.; Fishman, V.S.; Mazur, A.M.; Pomaznoy, M.; Khabarova, A.A.; Afonnikov, D.A.; Prokhortchouk, E.B.; Serov, O.L. Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. *Genome Biol.* **2015**, *16*, 77. [[CrossRef](#)]
46. Fraser, J.; Williamson, I.; Bickmore, W.A.; Dostie, J. An overview of genome organization and how we got there: From FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* **2015**, *79*, 347–372. [[CrossRef](#)] [[PubMed](#)]
47. Zhang, Z.; Li, G.; Toh, K.-C.; Sung, W.-K. 3D chromosome modeling with semi-definite programming and Hi-C data. *J. Comput. Biol.* **2013**, *20*, 831–846. [[CrossRef](#)] [[PubMed](#)]

48. Belaghzal, H.; Dekker, J.; Gibcus, J.H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **2017**, *123*, 56–65. [[CrossRef](#)] [[PubMed](#)]
49. Mifsud, B.; Tavares-Cadete, F.; Young, A.N.; Sugar, R.; Schoenfelder, S.; Ferreira, L.; Wingett, S.W.; Andrews, S.; Grey, W.; Ewels, P.A. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **2015**, *47*, 598. [[CrossRef](#)]
50. Liu, T.; Wang, Z. HiCNN: A very deep convolutional neural network to better enhance the resolution of Hi-C data. *Bioinformatics* **2019**, *35*, 4222–4228. [[CrossRef](#)] [[PubMed](#)]
51. Flyamer, I.M.; Gassler, J.; Imakaev, M.; Brandão, H.B.; Ulianov, S.V.; Abdennur, N.; Razin, S.V.; Mirny, L.A.; Tachibana-Konwalski, K. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **2017**, *544*, 110–114. [[CrossRef](#)] [[PubMed](#)]
52. Dong, Q.; Li, N.; Li, X.; Yuan, Z.; Xie, D.; Wang, X.; Li, J.; Yu, Y.; Wang, J.; Ding, B. Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice. *Plant J.* **2018**, *94*, 1141–1156. [[CrossRef](#)]
53. Li, J.; Zhang, W.; Li, X. 3D genome reconstruction with ShRec3D+ and Hi-C data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *15*, 460–468. [[CrossRef](#)]
54. Noble, W.; Duan, Z.-j.; Andronescu, M.; Schutz, K.; McIlwain, S.; Kim, Y.J.; Lee, C.; Shendure, J.; Fields, S.; Blau, C.A. A Three-Dimensional Model of the Yeast Genome. In *Proceedings of the International Conference on Research in Computational Molecular Biology, Vancouver, BC, Canada, 28–31 March 2011*; Springer: Berlin/Heidelberg, Germany, 2011; p. 320.
55. Baù, D.; Marti-Renom, M.A. Genome structure determination via 3C-based data integration by the Integrative Modeling Platform. *Methods* **2012**, *58*, 300–306. [[CrossRef](#)]
56. Rousseau, M.; Fraser, J.; Ferraiuolo, M.A.; Dostie, J.; Blanchette, M. Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinform.* **2011**, *12*, 414. [[CrossRef](#)]
57. Trussart, M.; Serra, F.; Baù, D.; Junier, I.; Serrano, L.; Marti-Renom, M.A. Assessing the limits of restraint-based 3D modeling of genomes and genomic domains. *Nucleic Acids Res.* **2015**, *43*, 3465–3477. [[CrossRef](#)]
58. Zhang, Z.; Li, G.; Toh, K.-C.; Sung, W.-K. Inference of Spatial Organizations of Chromosomes Using Semi-Definite Embedding Approach and Hi-C Data. In *Proceedings of the Annual International Conference on Research in Computational Molecular Biology, Padua, Italy, 10–13 May 2013*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 317–332.
59. Hu, M.; Deng, K.; Qin, Z.; Dixon, J.; Selvaraj, S.; Fang, J.; Ren, B.; Liu, J.S. Bayesian inference of spatial organizations of chromosomes. *PLoS Comput. Biol.* **2013**, *9*, e1002893. [[CrossRef](#)] [[PubMed](#)]
60. Varoquaux, N.; Ay, F.; Noble, W.S.; Vert, J.-P. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics* **2014**, *30*, i26–i33. [[CrossRef](#)] [[PubMed](#)]
61. Trieu, T.; Cheng, J. MOGEN: A tool for reconstructing 3D models of genomes from chromosomal conformation capturing data. *Bioinformatics* **2016**, *32*, 1286–1292. [[CrossRef](#)] [[PubMed](#)]
62. Rieber, L.; Mahony, S. miniMDS: 3D structural inference from high-resolution Hi-C data. *Bioinformatics* **2017**, *33*, i261–i266. [[CrossRef](#)]
63. Wlasnowolski, M.; Sadowski, M.; Czarnota, T.; Jodkowska, K.; Szalaj, P.; Tang, Z.; Ruan, Y.; Plewczynski, D. 3D-GNOME 2.0: A three-dimensional genome modeling engine for predicting structural variation-driven alterations of chromatin spatial structure in the human genome. *Nucleic Acids Res.* **2020**, *48*, W170–W176. [[CrossRef](#)]
64. Zhu, G.; Deng, W.; Hu, H.; Ma, R.; Zhang, S.; Yang, J.; Peng, J.; Kaplan, T.; Zeng, J. Reconstructing spatial organizations of chromosomes through manifold learning. *Nucleic Acids Res.* **2018**, *46*, e50. [[CrossRef](#)]
65. Abbas, A.; He, X.; Niu, J.; Zhou, B.; Zhu, G.; Ma, T.; Song, J.; Gao, J.; Zhang, M.Q.; Zeng, J. Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nat. Commun.* **2019**, *10*, 1–14. [[CrossRef](#)] [[PubMed](#)]
66. Caudai, C.; Salerno, E.; Zoppe, M.; Merelli, I.; Tonazzini, A. ChromStruct 4: A python code to estimate the chromatin structure from hi-C data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**, *16*, 1867–1878. [[CrossRef](#)] [[PubMed](#)]
67. Hua, N.; Tjong, H.; Shin, H.; Gong, K.; Zhou, X.J.; Alber, F. Producing genome structure populations with the dynamic and automated PGS software. *Nat. Protoc.* **2018**, *13*, 915. [[CrossRef](#)]
68. Nagano, T.; Lubling, Y.; Stevens, T.J.; Schoenfelder, S.; Yaffe, E.; Dean, W.; Laue, E.D.; Tanay, A.; Fraser, P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **2013**, *502*, 59–64. [[CrossRef](#)]
69. Bonev, B.; Cohen, N.M.; Szabo, Q.; Fritsch, L.; Papadopoulos, G.L.; Lubling, Y.; Xu, X.; Lv, X.; Hugnot, J.-P.; Tanay, A. Multiscale 3D genome rewiring during mouse neural development. *Cell* **2017**, *171*, 557–572.e24. [[CrossRef](#)]
70. Nagano, T.; Lubling, Y.; Várnai, C.; Dudley, C.; Leung, W.; Baran, Y.; Cohen, N.M.; Wingett, S.; Fraser, P.; Tanay, A. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **2017**, *547*, 61. [[CrossRef](#)] [[PubMed](#)]
71. Ramani, V.; Deng, X.; Qiu, R.; Gunderson, K.L.; Steemers, F.J.; Disteche, C.M.; Noble, W.S.; Duan, Z.; Shendure, J. Massively multiplex single-cell Hi-C. *Nat. Methods* **2017**, *14*, 263–266. [[CrossRef](#)]
72. Tan, L.; Xing, D.; Chang, C.-H.; Li, H.; Xie, X.S. Three-dimensional genome structures of single diploid human cells. *Science* **2018**, *361*, 924–928. [[CrossRef](#)] [[PubMed](#)]
73. Stevens, T.J.; Lando, D.; Basu, S.; Atkinson, L.P.; Cao, Y.; Lee, S.F.; Leeb, M.; Wohlfahrt, K.J.; Boucher, W.; O’Shaughnessy-Kirwan, A. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **2017**, *544*, 59–64. [[CrossRef](#)]
74. Carstens, S.; Nilges, M.; Habeck, M. Inferential structure determination of chromosomes from single-cell Hi-C data. *PLoS Comput. Biol.* **2016**, *12*, e1005292. [[CrossRef](#)]

75. Rosenthal, M.; Bryner, D.; Huffer, F.; Evans, S.; Srivastava, A.; Neretti, N. Bayesian estimation of three-dimensional chromosomal structure from single-cell hi-c data. *J. Comput. Biol.* **2019**, *26*, 1191–1202. [[CrossRef](#)] [[PubMed](#)]
76. Zhu, H.; Wang, Z. SCL: A lattice-based approach to infer 3D chromosome structures from single-cell Hi-C data. *Bioinformatics* **2019**, *35*, 3981–3988. [[CrossRef](#)]
77. Fudenberg, G.; Imakaev, M.; Lu, C.; Goloborodko, A.; Abdennur, N.; Mirny, L.A. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **2016**, *15*, 2038–2049. [[CrossRef](#)]
78. Sanborn, A.L.; Rao, S.S.; Huang, S.-C.; Durand, N.C.; Huntley, M.H.; Jewett, A.I.; Bochkov, I.D.; Chinnappan, D.; Cutkosky, A.; Li, J. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6456–E6465. [[CrossRef](#)]
79. Barbieri, M.; Chotalia, M.; Fraser, J.; Lavitas, L.-M.; Dostie, J.; Pombo, A.; Nicodemi, M. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16173–16178. [[CrossRef](#)] [[PubMed](#)]
80. Conte, M.; Fiorillo, L.; Bianco, S.; Chiariello, A.M.; Esposito, A.; Nicodemi, M. Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nat. Commun.* **2020**, *11*, 1–13. [[CrossRef](#)]
81. Jost, D.; Carrivain, P.; Cavalli, G.; Vaillant, C. Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* **2014**, *42*, 9553–9561. [[CrossRef](#)]
82. Di Pierro, M.; Zhang, B.; Aiden, E.L.; Wolynes, P.G.; Onuchic, J.N. Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12168–12173. [[CrossRef](#)]
83. Yu, N.; Polycarpou, A.A. Adhesive contact based on the Lennard–Jones potential: A correction to the value of the equilibrium distance as used in the potential. *J. Colloid Interface Sci.* **2004**, *278*, 428–435. [[CrossRef](#)] [[PubMed](#)]
84. Semiromi, D.T.; Azimian, A. Molecular dynamics simulation of nondroplets with the modified Lennard-Jones potential function. *Heat Mass Transf.* **2011**, *47*, 579–588. [[CrossRef](#)]
85. Smit, B. Phase diagrams of Lennard-Jones fluids. *J. Chem. Phys.* **1992**, *96*, 8639–8640. [[CrossRef](#)]
86. Mastny, E.A.; de Pablo, J.J. Melting line of the Lennard-Jones system, infinite size, and full potential. *J. Chem. Phys.* **2007**, *127*, 104504. [[CrossRef](#)]
87. Wei-Zhong, L.; Cong, C.; Jian, Y. Molecular dynamics simulation of self-diffusion coefficient and its relation with temperature using simple Lennard-Jones potential. *Heat Transf.* **2008**, *37*, 86–93. [[CrossRef](#)]
88. Pandey, R.B.; Farmer, B.L. Conformational response to solvent interaction and temperature of a protein (histone h3. 1) by a multi-grained Monte Carlo simulation. *PLoS ONE* **2013**, *8*. [[CrossRef](#)] [[PubMed](#)]
89. Duan, Z.; Andronescu, M.; Schutz, K.; McIlwain, S.; Kim, Y.J.; Lee, C.; Shendure, J.; Fields, S.; Blau, C.A.; Noble, W.S. A three-dimensional model of the yeast genome. *Nature* **2010**, *465*, 363–367. [[CrossRef](#)] [[PubMed](#)]
90. Lieberman-Aiden, E.; Van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragozy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293. [[CrossRef](#)]
91. Engreitz, J.M.; Pandya-Jones, A.; McDonel, P.; Shishkin, A.; Sirokman, K.; Surka, C.; Kadri, S.; Xing, J.; Goren, A.; Lander, E.S. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **2013**, *341*, 1237973. [[CrossRef](#)] [[PubMed](#)]
92. Beagrie, R.A.; Scialdone, A.; Schueler, M.; Kraemer, D.C.; Chotalia, M.; Xie, S.Q.; Barbieri, M.; de Santiago, I.; Lavitas, L.-M.; Branco, M.R. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **2017**, *543*, 519–524. [[CrossRef](#)] [[PubMed](#)]
93. Binder, K. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*; Oxford University Press: Oxford, UK, 1995.
94. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [[CrossRef](#)] [[PubMed](#)]
95. Zhang, Y.; Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct. Funct. Bioinform.* **2004**, *57*, 702–710. [[CrossRef](#)]