Supplementary Materials

January 2, 2018

Supplementary Methods

Manifold learning framework

Recently, manifold learning, such as t-SNE [1], has been successfully applied as a general framework for nonlinear dimensionality reduction in machine learning and pattern recognition [2, 3, 4, 5]. It aims to reconstruct the underlying low-dimensional manifolds from the abstract representations in the high-dimensional space, that is, it uncovers the intrinsic low-dimensional manifolds which preserve the local neighbourhoods of high-dimensional data (Supplementary Figure S1).

In practice, many high-dimensional data of interest lie on the structures that are intrinsically embedded from a low-dimensional manifold. For example, in computer vision, images of faces can be regarded as points in a high-dimensional vector space, in which each dimension corresponds to the brightness of every pixel in the image. All the images can be considered to lie on an intrinsic 3D manifold parameterized by two pose variables and and an azimuthal lighting orientation angle [2]. In our chromatin structure modeling problem, the meaningful spatial organizations of chromosomes can be interpreted as the geometry of manifolds in 3D Euclidean space. The Hi-C interaction frequency data can be regarded as a specific representation of the neighboring affinities reflecting the spatial arrangements of genomic loci, which is intrinsically determined by the underlying manifolds embedded in Hi-C space. Thus, manifold learning can be used here to uncover the meaningful geometry of manifolds in low-dimensional space based on a process of neighborhood embedding, which preserves the local neighborhood of genomic loci in Hi-C space.

Modeling the conformational energy of chromatin structures

According to our known biophysical knowledge of a polymer model [6, 7, 8], the physical potential of a chromatin conformation can be described by an energy function $E^{(m)}$ consisting of three terms, including the stretching energy E_{stretch}, the bending energy E_{bend} and the excluding energy E_{exclude}, that is,

$$E^{(m)} = E^{(m)}_{\text{stretch}} + E^{(m)}_{\text{bend}} + E^{(m)}_{\text{exclude}}. \tag{1}$$

First, the stretching energy term E_{stretch} accounts for the stretching resistance of chromatin fibers, which is defined by the following equation

$$E^{(m)}_{\text{stretch}} = \sum_{i=1}^{n-1} \frac{1}{2} k_s (\|s^{(m)}_{i+1} - s^{(m)}_i\| - l_i)^2, \tag{2}$$

where k_s stands for the bond spring constant characterizing the chromatin stiffness, and l_i stands for the equilibrium length of the i-th segment in the modeled chromatin structure. Since the sequence length for each pair of adjacent genomic loci is known, their corresponding distance l_i can be generally derived based on the packing density, which is usually assigned with 130bp/nm [9, 10] in approximate 10 kb resolution and can be computed based on a 1/3 power-law relationship between spatial distances and corresponding genomic distances in other resolutions (which can be derived mainly based on the 3D FISH data [11] or the fractal globule model [12]).

Second, the bending energy term E_{bend} accounts for the bending potential of chromatin fibers [13], which is defined by the following equation,

$$E_{\text{bend}}^{(m)} = \sum_{i=2}^{n-1} \frac{1}{2} k_\theta \left\langle s_{i+1}^{(m)} - s_i^{(m)}, s_i^{(m)} - s_{i-1}^{(m)} \right\rangle^2, \tag{3}$$

where k_θ denotes the bending energy constant and $\langle \cdot \rangle$ denotes the angle between two adjacent segments.

Third, the excluding energy term E_{stretch} accounts for the inter-particle repulsive potential. It takes the form of the repulsive part of the Lennard-Jones potential [14, 15], that is,

$$E_{\text{exclude}}^{(m)} = \begin{cases} \displaystyle\sum_{2 \le i+1 \le j \le n} 4\varepsilon \left[\left(\frac{d_0}{d_{ij}^{(m)}} \right)^{12} - \left(\frac{d_0}{d_{ij}^{(m)}} \right)^{6} + \frac{1}{4} \right], & d_{ij}^{(m)} < 2^{\frac{1}{6}} d_0; \\ 0, \quad \text{otherwise}, \end{cases} \tag{4}$$

where $d_{ij}^{(m)} = \| s_i^{(m)} - s_j^{(m)} \|$ stands for the Euclidean distance between $s_i^{(m)}$ and $s_j^{(m)}$, ε is a Lennard-Jones energy parameter, and d_0 is the Lennard-Jones size parameter, which represents the distance threshold within which the repulsive force is zero. For the parameters in $E^{(m)}$, we use the same setting as in [15, 16]. The details of these parameter choices are provided in Supplementary Table S1.

Optimization of the objective function

The overall cost function C of GEM is defined as

$$C = C_1 + \lambda_E C_2, \tag{5}$$

We use gradient descent to minimize C in Equation (5). In particular, the gradient of C is calculated as follows

$$\frac{\partial C}{\partial s_i^{(m)}} = \frac{\partial C_1}{\partial s_i^{(m)}} + \lambda_E \pi^{(m)} \frac{\partial E^{(m)}}{\partial s_i^{(m)}}, \tag{6}$$

$$\frac{\partial C}{\partial \pi^{(m)}} = \frac{\partial C_1}{\partial \pi^{(m)}} + \lambda_E E^{(m)}. \tag{7}$$

The gradient of C_1 with respect to the coordinates $s_i^{(m)}$ in 3D Euclidean space is given by

$$\frac{\partial C_1}{\partial s_i^{(m)}} = 4 \sum_j \frac{\partial C_1}{\partial d_{ij}^{(m)}} (s_i^{(m)} - s_j^{(m)}), \tag{8}$$

where $d_{ij}^{(m)}$ stands for the Euclidean distance between $s_i^{(m)}$ and $s_j^{(m)}$, and

$$\frac{\partial C_1}{\partial d_{ij}^{(m)}} = \frac{\pi^{(m)}(p_{ij} - q_{ij})(1 + d_{ij}^{(m)})^{-2}}{q_{ij}Z}, \tag{9}$$

$$Z = \sum_k \sum_{l \neq k} \sum_{m'} \pi^{(m')}(1 + \|s_k^{(m')} - s_l^{(m')}\|^2)^{-1}. \tag{10}$$

From the term $(1+d_{ij}^{(m)})^{-2}$ in the above gradient, we can also see that two neighboring nodes are not likely to be modeled by widely separated points. Also, $\pi^{(m)}$ is defined as follows to satisfy the definition of probability, that is,

$$\pi^{(m)} = \frac{e^{-w^{(m)}}}{\sum_{m'} e^{-w^{(m')}}}. \tag{11}$$

Then we can obtain a new version of the gradient

$$\frac{\partial C_1}{\partial w^{(m)}} = \pi^{(m)} \left(\left(\sum_{m'} \pi^{(m')} \frac{\partial C_1}{\partial \pi^{(m')}} \right) - \frac{\partial C_1}{\partial \pi^{(m)}} \right), \tag{12}$$

where

$$\frac{\partial C_1}{\partial \pi^{(m)}} = \sum_k \sum_{l \neq k} \frac{(q_{kl} - p_{kl})(1 + d_{kl}^{(m)})^{-1}}{q_{kl}Z}. \tag{13}$$

The gradient of $E^{(m)}$ with respect to $s_i^{(m)}$ includes three parts, that is,

$$\frac{\partial E^{(m)}}{\partial s_i^{(m)}} = \frac{\partial E_{\text{stretch}}^{(m)}}{\partial s_i^{(m)}} + \frac{\partial E_{\text{bend}}^{(m)}}{\partial s_i^{(m)}} + \frac{\partial E_{\text{exclude}}^{(m)}}{\partial s_i^{(m)}}. \tag{14}$$

The first part is given by

$$\frac{\partial E_{\text{stretch}}^{(m)}}{\partial s_i^{(m)}} = \begin{cases} (G_{11} + G_{12}), & i = 2, \cdots, n-1; \\ G_{11}, & i = 1; \\ G_{12}, & i = n, \end{cases} \tag{15}$$

where

$$G_{11} = -k_s(\|\Delta s_i^{(m)}\| - l_i)\frac{\Delta s_i^{(m)}}{\|\Delta s_i^{(m)}\|}, \tag{16}$$

$$G_{12} = k_s(\|\Delta s_{i-1}^{(m)}\| - l_{i-1})\frac{\Delta s_{i-1}^{(m)}}{\|\Delta s_{i-1}^{(m)}\|}, \tag{17}$$

$$\Delta s_i^{(m)} = s_{i+1}^{(m)} - s_i^{(m)}. \tag{18}$$

The second part is given by

$$\frac{\partial E_{\text{bend}}^{(m)}}{\partial s_i^{(m)}} = \begin{cases} (G_{21} + G_{22} + G_{23}), & i = 3, \cdots, n-2; \\ G_{23}, & i = 1; \\ G_{21} + G_{23}, & i = 2; \\ G_{21} + G_{22}, & i = n-1; \\ G_{22}, & i = n, \end{cases} \tag{19}$$

where

$$G_{21} = \frac{k_\theta \arccos \frac{\Delta s_i^{(m)} \Delta s_{i-1}^{(m)}}{\|\Delta s_i^{(m)}\| \|\Delta s_{i-1}^{(m)}\|}}{\sqrt{\|\Delta s_i^{(m)}\|^2 \|\Delta s_{i-1}^{(m)}\|^2 - (\Delta s_i^{(m)} \Delta s_{i-1}^{(m)})^2}} \cdot \left(\Delta s_i^{(m)} - \Delta s_{i-1}^{(m)} + \frac{\Delta s_i^{(m)} \Delta s_{i-1}^{(m)}}{\|\Delta s_i^{(m)}\|^2} \Delta s_i^{(m)} - \frac{\Delta s_i^{(m)} \Delta s_{i-1}^{(m)}}{\|\Delta s_{i-1}^{(m)}\|^2} \Delta s_{i-1}^{(m)} \right), \tag{20}$$

$$G_{22} = \frac{k_\theta \arccos \frac{\Delta s_{i-1}^{(m)} \Delta s_{i-2}^{(m)}}{\|\Delta s_{i-1}^{(m)}\| \|\Delta s_{i-2}^{(m)}\|}}{\sqrt{\|\Delta s_{i-1}^{(m)}\|^2 \|\Delta s_{i-2}^{(m)}\|^2 - (\Delta s_{i-1}^{(m)} \Delta s_{i-2}^{(m)})^2}} \left(\Delta s_{i-2}^{(m)} - \frac{\Delta s_{i-1}^{(m)} \Delta s_{i-2}^{(m)}}{\|\Delta s_{i-1}^{(m)}\|^2} \Delta s_{i-1}^{(m)} \right), \tag{21}$$

$$G_{23} = \frac{k_\theta \arccos \frac{\Delta s_{i+1}^{(m)} \Delta s_i^{(m)}}{\|\Delta s_{i+1}^{(m)}\| \|\Delta s_i^{(m)}\|}}{\sqrt{\|\Delta s_{i+1}^{(m)}\|^2 \|\Delta s_i^{(m)}\|^2 - (\Delta s_{i+1}^{(m)} \Delta s_i^{(m)})^2}} \left(-\Delta s_{i+1}^{(m)} + \frac{\Delta s_{i+1}^{(m)} \Delta s_i^{(m)}}{\|\Delta s_i^{(m)}\|^2} \Delta s_i^{(m)} \right). \tag{22}$$

The third part is given by

$$\frac{\partial E_{\text{exclude}}^{(m)}}{\partial s_i^{(m)}} = \begin{cases} \sum_{1 \le j \le n} 24\varepsilon \left(-2 d_0^{12} \frac{s_i^{(m)} - s_j^{(m)}}{{d_{ij}^{(m)}}^{14}} + d_0^6 \frac{s_i^{(m)} - s_j^{(m)}}{{d_{ij}^{(m)}}^8} \right), & d_{ij}^{(m)} < 2^{\frac{1}{6}} d_0 \quad \& \quad i \neq j; \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

Based on the above derivations, we can develop an adaptive gradient descent method to solve the optimization problem in Equation (5). In particular, the learning rates for $\pi^{(m)}$ and $s_i^{(m)}$ can be changed constantly during different stages to accelerate the optimization process. However, there may exist a "sinking" problem in such an optimization strategy, if the random initialization process leads to a huge difference of goodness between a pair of conformations. In such a case, if the learning rate for $\pi^{(m)}$ is too large, $\pi^{(m)}$ of those structures that deviated largely from random initialization will sink at zero rapidly, and these conformations will not be considered during the downstream optimization process, mainly due to the vanishing gradient, although they can still converge to proper solutions if a sufficient number of optimization iterations are performed.

To address this problem, we could employ two strategies during the optimization process, i.e., two-stage optimization and asynchronous starting. The first strategy is to divide the whole optimization procedure into two stages, including average-structure optimization and multi-conformation optimization. In particular, we first compute an average structure using a single-conformation version of GEM. Then, this average structure is used as an initial structure for the second-stage optimization, which is accomplished through a multi-conformation version of GEM. Initialization from such a pre-computed structure that is not so far away the final solution provides a beneficial guidance for optimization. More importantly, the goodness scores of the initial conformations in the second-stage optimization do not has extremely large variance, which thus can prevent the aforementioned sinking problem. The second strategy that we could use is to delay the update of the learning rates of $\pi^{(m)}$ in the second-stage optimization if the goodness scores of the initial conformations remain considerably different. Under such a strategy, it is unlikely that the sinking problem will occur, after a number of iteration steps to optimize the conformations with relatively fixed weights. In practice, we found that the first strategy is often sufficient enough to prevent the sinking problem.

Generation of simulated Hi-C data

The simulated Hi-C data were generated according to the following procedure. At the beginning, we applied the Brownian simulation method to generate a set of random chromatin conformations, each of which imitated a real chromatin conformation in Hi-C experiments. Let N_c denote the total number of cells. Then we mimicked the experimental Hi-C protocol to obtain the simulated Hi-C interaction frequency map. First, the restricted sites along the synthetic chromatin conformations were chosen randomly and then cleaved by the restriction enzymes. Second, as in the study by Trussart et al. [17], we used a Gaussian process to generate the genomic interactions between restriction sites. In particular, let P_m and σ denote the scaling factor (i.e., the maximum interaction probability) and the standard deviation of the Gaussian function describing the relationship between the probability of generating the interaction and the spatial distance between a pair of genomic loci. Let d_{interact} denote the spatial distance between a pair of genomic loci and P_{interact} denote the probability of generating the corresponding genomic interaction. The Gaussian process of generating the genomic interaction can be described as,

$$P_{\text{interact}} = P_m e^{(-\frac{d^2_{\text{interact}}}{2\sigma^2})}. \tag{24}$$

Due to experimental uncertainty, not all the interactions between restriction sites can be captured by Hi-C experiments. Here, we used the trapping rate α_t to model such experimental uncertainty, this is, with probability α_t, an occurred interaction is observed between restriction sites, otherwise it is missed with probability $1 - \alpha_t$. After considering the trapping rate α_t, Equation (24) can be rewritten into the following form,

$$P_{\text{interact}} = \begin{cases} P_m e^{(-\frac{d^2_{\text{interact}}}{2\sigma^2})}, & \text{with probability } \alpha_t; \\ 0, & \text{with probability } 1 - \alpha_t. \end{cases} \tag{25}$$

Overall, the simulation process for generating a synthetic interaction frequency dataset can be determined by parameters $(\alpha_t, P_m, \sigma, N_c)$.

The 10-fold cross-validation procedure

The Hi-C data of a chromosome were randomly divided into 10 roughly equal-sized subsets. Nine of them were selected as training data and input into GEM to compute the chromatin structures. Based on the latent function between interaction frequencies and spatial distances between genomic loci derived by GEM (see the dashed box in Figure 1), the modeled chromatin structures can also be used to obtain the reconstructed or predicted Hi-C map. Then the remaining subset was held as test data to assess the accuracy of the modeled conformations by comparing the original Hi-C map to the reconstructed Hi-C map. Such a process was performed 10 folds, and the average result was used to evaluate the final modeling performance.

The convergence and parameter selection of GEM

We examined the convergence of the optimization procedure employed in GEM, which is a two-stage optimization scheme including average-structure optimization and multi-conformation optimization. As shown in Supplementary Figure S6a, both optimization stages converged successfully. In the first optimization stage, which aimed at computing the average structure, the cost function descended rapidly at the beginning. After 2000 iterations, the cost function

began to converge, indicating a stable average structure was reached. After that, the second optimization stage was performed to obtain multiple conformations. Probably because the average structure had been determined, the cost function only descended slightly in this stage. Overall, the second optimization stage converged after approximate 20000 iterations.

In our model, the only parameter, the coefficient of the energy term λ_E, determines a trade-off between the fitness of the spatial constraint derived from Hi-C data and structural feasibility measured in terms of conformational energy. The parameter λ_E can be decided by the users according to their emphasized aspects. Alternatively, this parameter can be determined by the following two automatic methods.

First, inspired by previous studies [18, 16, 19], we can use a Bayesian approach to determine the proper value of the coefficient of the energy term, λ_E, based on extra priori knowledge, such as the volume of a chromosome obtained by direct experimental observations or estimated by indirect experimental observations (e.g., DNA density [20]). The goal is to select a proper value of λ_E that best interprets both the input Hi-C data and the observation about the volume of a chromosome. The posterior probability $\Pr(\lambda_E|H, V)$ of the coefficient λ_E given Hi-C data H and the volume of a chromosome V can be derived according to Bayes' theorem, i.e.,

$$\begin{aligned}\Pr(\lambda_E|H, V) &= \frac{\Pr(H, V|\lambda_E)\Pr(\lambda_E)}{\Pr(H, V)} \\ &= \frac{\Pr(H|\lambda_E)\Pr(V|\lambda_E)\Pr(\lambda_E)}{\Pr(H)\Pr(V)} \\ &\propto \Pr(H|\lambda_E)\Pr(V|\lambda_E)\Pr(\lambda_E).\end{aligned} \tag{26}$$

Based on the maximum a posteriori estimation of λ_E, we define a Bayesian score to evaluate the parameter λ_E for our model. We assume that the prior distribution of λ_E is uniform, and thus $\Pr(\lambda_E)$ can be considered constant and is not necessary to be included in the Bayesian score. The final optimized value of the KL divergence C_1 mentioned in Equation (5), which is dependent on λ_E during the optimization process, measures the degree of mismatch (ranging from 0 to 1) between structures calculated with λ_E and the Hi-C data. Here, we use $1 - C_1$ to define $\Pr(H|\lambda_E)$. In addition, the mismatch between the volume of a computed structure v' (which is also dependent on λ_E) and its real volume v can be measured by the relative error ratio. Here, we use the inverse of this relative error ratio to define $\Pr(V|\lambda_E)$. To sum up, the Bayesian score is defined as,

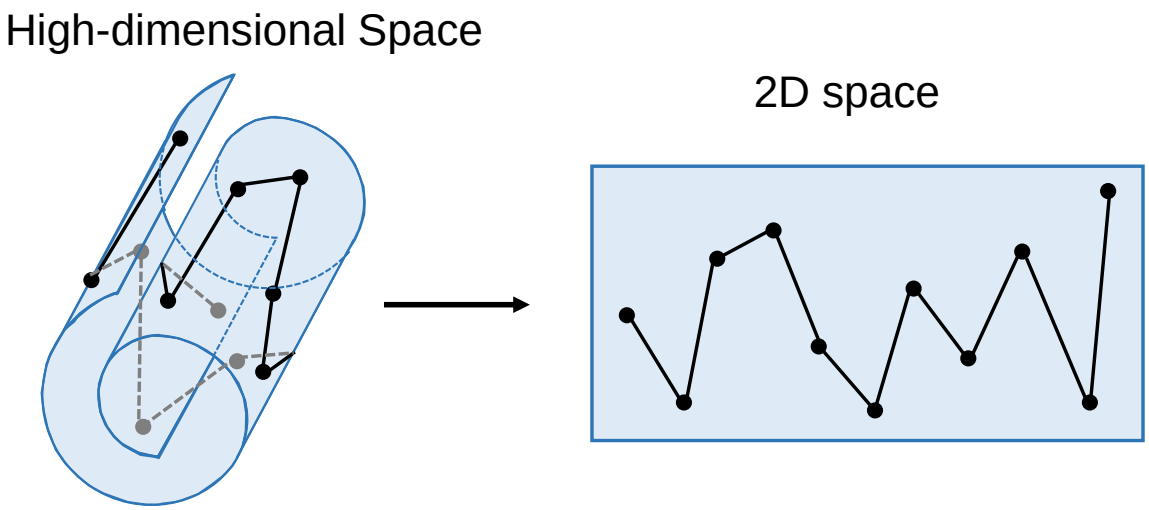
$$\begin{aligned}\textbf{Bayesian score} &= \Pr(\lambda_E|H, V) \\ &= \Pr(H|\lambda_E)\Pr(V|\lambda_E)\Pr(\lambda_E) \\ &\propto (1 - C_1)(\frac{v}{|v - v'|}).\end{aligned} \tag{27}$$

Second, if we do not have any priori knowledge, we can transform the parameter selection into a multi-criteria decision problem and use TOPSIS [21] to obtain the best estimate of parameter λ_E. In this setting, each value of different λ_E is regarded as a decision that is evaluated by two criteria C_1 and C_2.

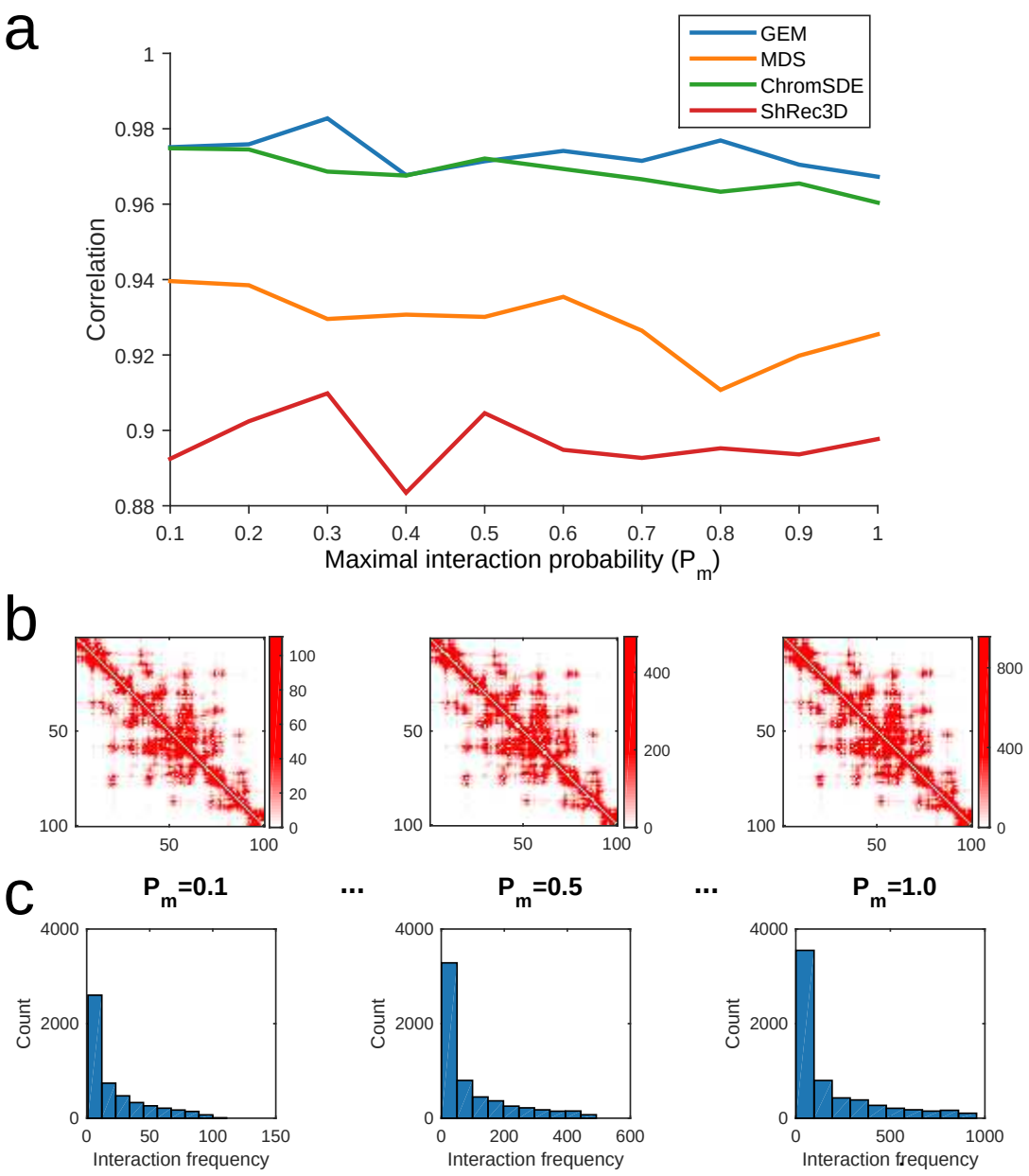
In our study, we employed the Bayesian approach to perform parameter selection. We computed the Bayesian scores with respect to a wide range of the coefficient of energy term λ_E (Supplementary Figure S6b) and then chose $\lambda_E = 5 \times 10^{-12}$ which had the maximum Bayesian score, as the most reasonable parameter for human chromosome 14 (which is marked by the orange dashed line in Supplementary Figure S6b). In practice, we only need to select a rough range for λ_E because in general the change of λ_E within the same order of magnitude has little influence on the performance of our model, measured by the Pearson correlation between

experimental and reconstructed Hi-C data in the 10-fold cross-validation procedure (Supplementary Figure S6c).

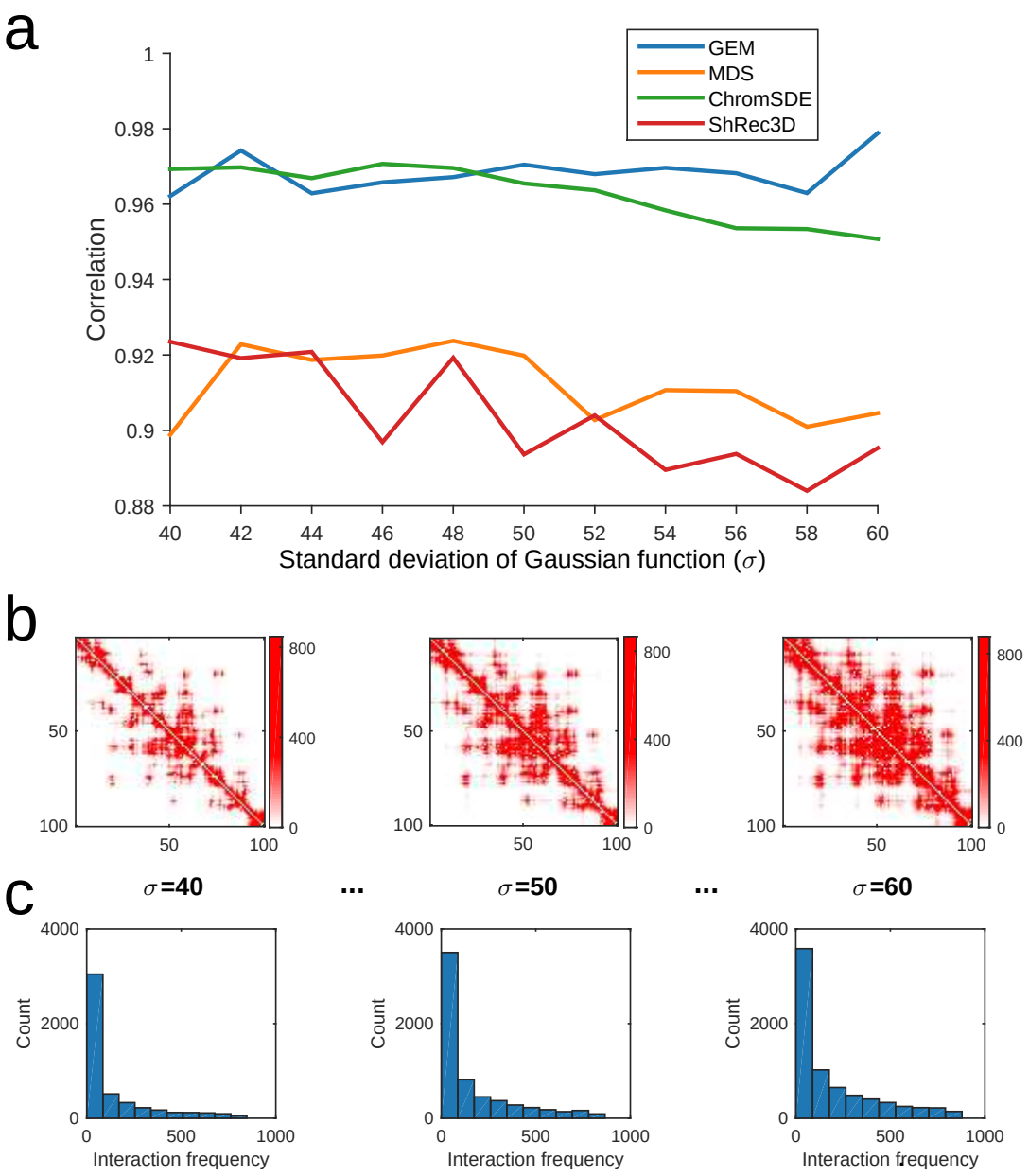
Supplementary Figures and Tables



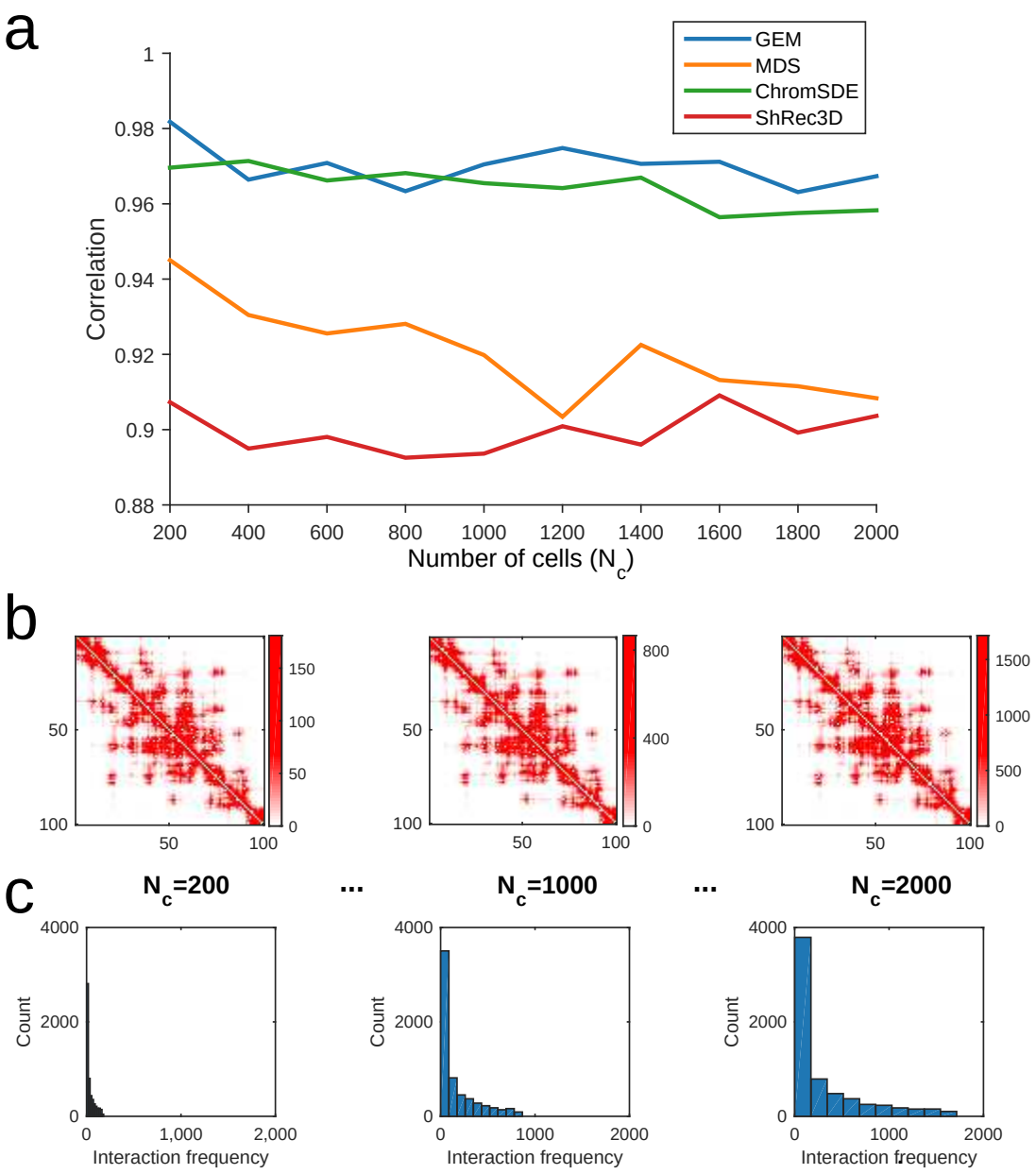
Supplementary Figure S1: A schematic illustration of manifold learning. In high-dimensional (3D for example) space, the data points lie on a "Swiss roll" structure. After embedding the data points from high-dimensional space into low-dimensional (2D for example) space by manifold learning, the intrinsic manifold is uncovered.



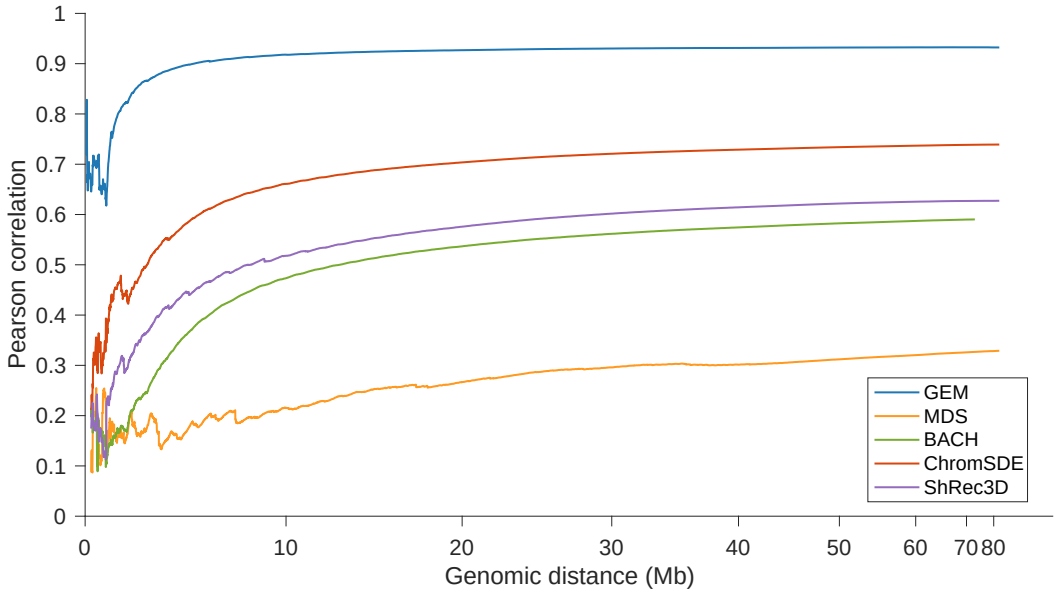
Supplementary Figure S2: The validation results on the simulated Hi-C data, which were generated according to different settings of the maximum interaction probability P_m (see Supplementary Methods). **(a)** The comparisons of Pearson correlations between GEM and other modeling methods, including the MDS [22, 23] based model, ChromSDE [24] and ShRec3D [25]. **(b)** and **(c)** show the typical examples of the simulated Hi-C maps and the corresponding distributions of the simulated interaction frequencies as P_m increases, respectively. In the simulated Hi-C maps, the axes denote the genomic loci (1 Mb resolution) and the values of the entries indicate the simulated interaction frequencies. In the histograms, the x axes denote the interaction frequencies obtained from the Hi-C maps and the y axes denote the numbers of data points falling into individual interaction frequency intervals.



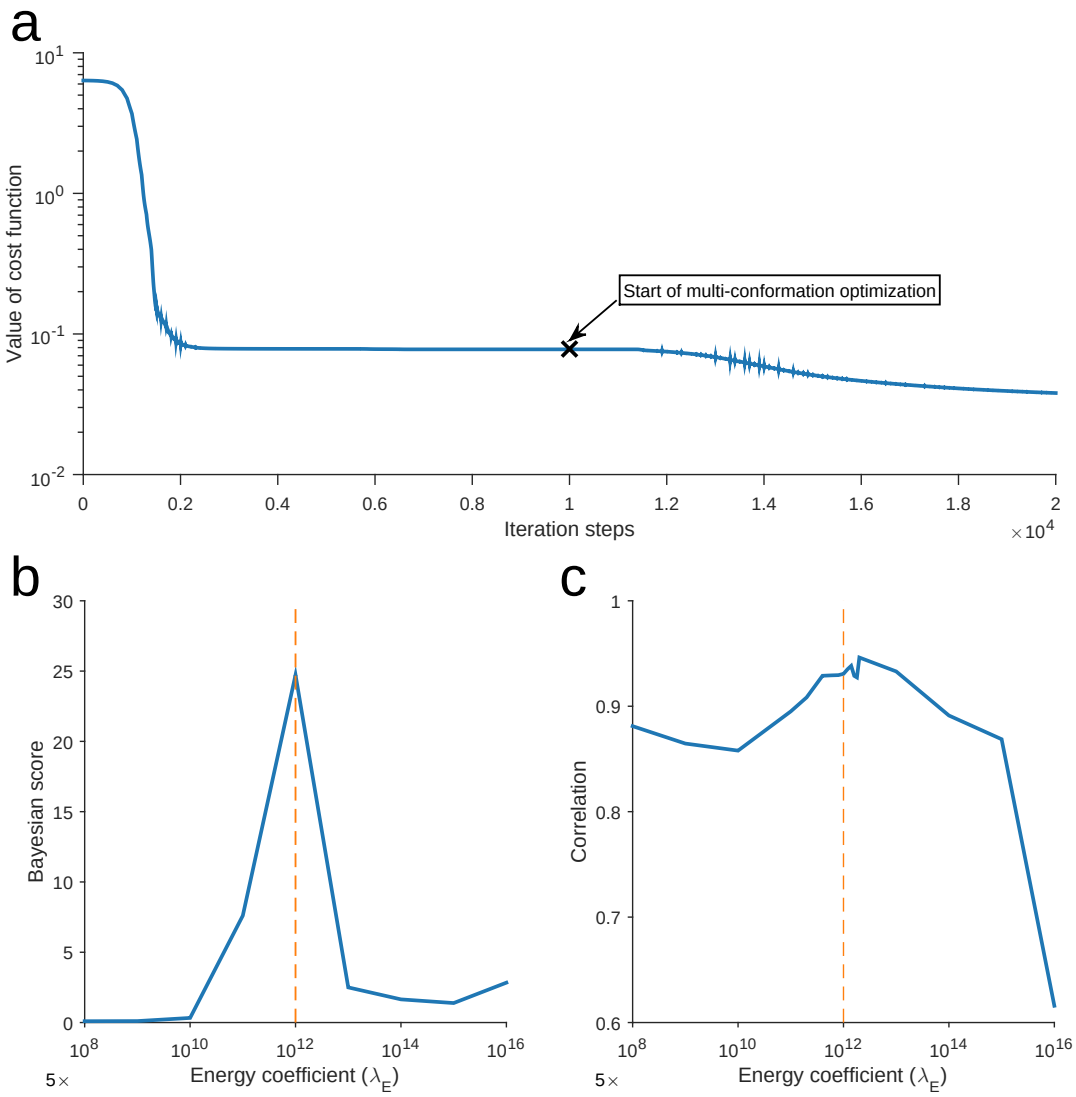
Supplementary Figure S3: The validation results on the simulated Hi-C data, which were generated according to different settings of the standard deviation of Gaussian function σ (see Supplementary Methods). **(a)** The comparisons of Pearson correlations between GEM and other modeling methods, including the MDS [22, 23] based model, ChromSDE [24] and ShRec3D [25]. **(b)** and **(c)** show the typical examples of the simulated Hi-C maps and the corresponding distributions of the simulated interaction frequencies as σ increases, respectively. In the simulated Hi-C maps, the axes denote the genomic loci (1 Mb resolution) and the values of the entries indicate the simulated interaction frequencies. In the histograms, the x axes denote the interaction frequencies obtained from the Hi-C maps and the y axes denote the numbers of data points falling into individual interaction frequency intervals.



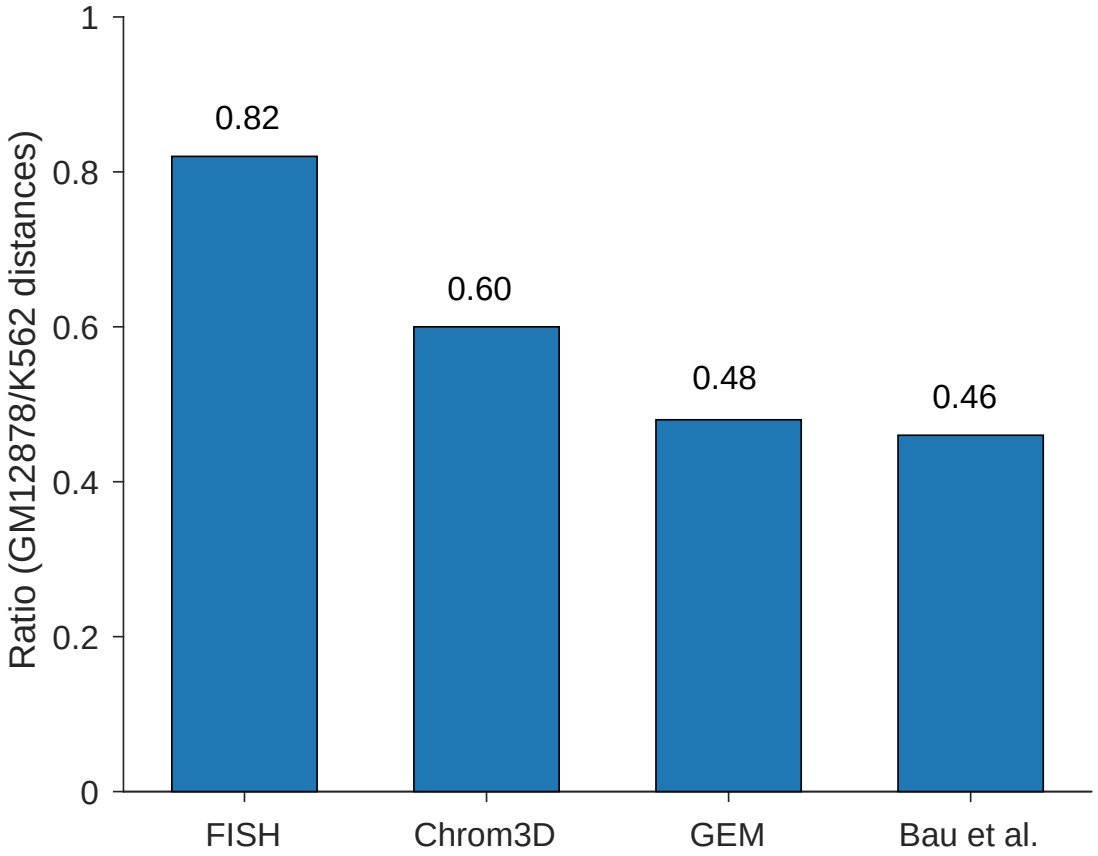
Supplementary Figure S4: The validation results on the simulated Hi-C data, which were generated according to different settings of the number of cells N_c (see Supplementary Methods). **(a)** The comparisons of Pearson correlations between GEM and other modeling methods, including the MDS [22, 23] based model, ChromSDE [24] and ShRec3D [25]. **(b)** and **(c)** show the typical examples of the simulated Hi-C maps and the corresponding distributions of the simulated interaction frequencies as N_c increases, respectively. In the simulated Hi-C maps, the axes denote the genomic loci (1 Mb resolution) and the values of the entries indicate the simulated interaction frequencies. In the histograms, the x axes denote the interaction frequencies obtained from the Hi-C maps and the y axes denote the numbers of data points falling into individual interaction frequency intervals.



Supplementary Figure S5: Comparison results between different modeling methods, showing the plot of the Pearson correlations between experimental Hi-C data and predicted interaction frequencies of human chromosome 14 as a function of genomic distance.



Supplementary Figure S6: The convergence and parameter selection of GEM. Here, human chromosome 14 at a resolution of 1 Mb was used as an example. **(a)** The convergence results on the cost function C in Equation (5). The first-stage optimization (i.e., average-structure optimization) took about 1000 iterations and stopped at the position marked with black cross, which denotes the start of the second-stage optimization (i.e., multi-conformation optimization), which also took about 1000 iterations. **(b)** The Bayesian score as a function of the coefficient parameter λ_E that weighs the conformational energy term. The value of λ_E (5×10^{-12}) with the maximum Bayesian score was used in GEM, which is marked by the orange dashed line. **(c)** The 10-fold cross-validation results (evaluated in terms of the Pearson correlation between experimental and reconstructed Hi-C data) as a function of the coefficient parameter λ_E for the conformational energy term. The orange dashed line indicates the value of λ_E used in GEM.



Supplementary Figure S7: Comparisons between different modeling methods on the ENm008 ENCODE region containing the α-globin locus in terms of the ratios of spatial distances between two modeled regions (i.e., positions 55911-56690 and 402437-418222 on chromosome 16) between GM12878 and K562 cells. The ratio of the corresponding spatial distances derived from the FISH probes in [26] is also shown.

Supplementary Table S1: The parameter setting of the conformational energy model. k_B denotes the Boltzmann constant and T denotes the absolute temperature. T is set to 300K according to [15, 16].

Parameter	Symbol	Reduced unit	SI unit
Bond spring constant	k_s	$500k_BT/d_0^2$	$2.3 \times 10^{-3} J \cdot m^{-2}$
Bending energy constant	k_θ	$4k_BT/rad^2$	$1.7 \times 10^{-20} J \cdot rad^{-2}$
Lennard-Jones energy parameter	ε	$1.0k_BT$	$4.1 \times 10^{-21} J$
Lennard-Jones size parameter	d_0	1.0	$30nm$

Supplementary Table S2: The enrichments of the functional elements (compared with the background) in both missing and known chromatin loops on human chromosome 19.

	Missing loops	Known loops
H3K27ac	1.4020	1.1203
DNase-seq	1.7797	1.6842
H3K4me3	1.4423	1.3410
H3K4me1	1.2934	1.1034
Promoters	1.5563	2.4234
Enhancers	3.1685	2.5611

References

[1] Maaten, L. v. d. and Hinton, G. (2008) Journal of Machine Learning Research **9(Nov)**, 2579–2605.

[2] Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000) science **290(5500)**, 2319–2323.

[3] Borg, I. and Groenen, P. J. (2005) Modern multidimensional scaling: Theory and applications, Springer Science & Business Media, .

[4] Lee, J. A. and Verleysen, M. (2007) Nonlinear dimensionality reduction, Springer Science & Business Media, .

[5] Lawrence, N. D. (2012) Journal of Machine Learning Research **13(May)**, 1609–1638.

[6] Münkel, C. and Langowski, J. (1998) Physical Review E **57(5)**, 5888.

[7] Mateos-Langerak, J., Bohn, M., deLeeuw, W., Giromus, O., Manders, E. M., Verschure, P. J., Indemans, M. H., Gierman, H. J., Heermann, D. W., Van Driel, R., et al. (2009) Proceedings of the National Academy of Sciences **106(10)**, 3812–3817.

[8] Schiessel, H. (2003) Journal of Physics: Condensed Matter **15(19)**, R699.

[9] Jhunjhunwala, S., vanZelm, M. C., Peak, M. M., Cutchin, S., Riblet, R., vanDongen, J. J., Grosveld, F. G., Knoch, T. A., and Murre, C. (2008) Cell **133(2)**, 265–279.

[10] Bystricky, K., Heun, P., Gehlen, L., Langowski, J., and Gasser, S. M. (2004) Proceedings of the National Academy of Sciences of the United States of America **101(47)**, 16495–16500.

[11] Varoquaux, N., Ay, F., Noble, W. S., and Vert, J.-P. (2014) Bioinformatics **30(12)**, i26–i33.

[12] Mirny, L. A. (2011) Chromosome research **19(1)**, 37–51.

[13] Langowski, J. and Heermann, D. W. (2007) In Seminars in cell & developmental biology volume **18**, Elsevier : pp. 659–667.

[14] Verlet, L. (1967) Physical review **159(1)**, 98.

[15] Meluzzi, D. and Arya, G. (2012) Nucleic acids research p. gks1029.

[16] Wang, S., Xu, J., and Zeng, J. (2015) Nucleic acids research **43(8)**, e54–e54.

[17] Trussart, M., Serra, F., Baù, D., Junier, I., Serrano, L., and Marti-Renom, M. A. (2015) Nucleic acids research **43(7)**, 3465–3477.

[18] Rieping, W., Habeck, M., and Nilges, M. (2005) Science **309(5732)**, 303–306.

[19] Carstens, S., Nilges, M., and Habeck, M. (2016) PLOS Computational Biology **12(12)**, e1005292.

[20] Rosa, A. and Everaers, R. (2008) PLoS Comput Biol **4(8)**, e1000153.

[21] Hwang, C. and Yoon, K. (1981) New York (1981).

[22] Torgerson, W. S. (1952) Psychometrika **17(4)**, 401–419.

[23] Young, G. and Householder, A. S. (1938) Psychometrika **3(1)**, 19–22.

[24] Zhang, Z., Li, G., Toh, K.-C., and Sung, W.-K. (2013) Journal of computational biology **20(11)**, 831–846.

[25] Lesne, A., Riposo, J., Roger, P., Cournac, A., and Mozziconacci, J. (2014) Nature methods **11(11)**, 1141–1143.

[26] Baù, D., Sanyal, A., Lajoie, B. R., Capriotti, E., Byron, M., Lawrence, J. B., Dekker, J., and Marti-Renom, M. A. (2011) Nature structural & molecular biology **18(1)**, 107–114.