*Published online 15 September 2017* *Nucleic Acids Research, 2017, Vol. 45, No. 20* ***11547–11558***
*doi: 10.1093/nar/gkx784*

# Computational construction of 3D chromatin ensembles and prediction of functional interactions of alpha-globin locus from 5C data

**Gamze Gürsoy[1,†], Yun Xu[1,†], Amy L. Kenter[2] and Jie Liang[1,*]**

[1]Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60607, USA and [2]Department of Microbiology and Immunology, University of Illinois College of Medicine, Chicago, IL 60612, USA

Received December 07, 2016; Revised August 23, 2017; Editorial Decision August 24, 2017; Accepted August 30, 2017

## ABSTRACT

**Conformation capture technologies measure frequencies of interactions between chromatin regions. However, understanding gene-regulation require knowledge of detailed spatial structures of heterogeneous chromatin in cells. Here we describe the nC-SAC (n-Constrained-Self Avoiding Chromatin) method that transforms experimental interaction frequencies into 3D ensembles of chromatin chains. nC-SAC first distinguishes specific from non-specific interaction frequencies, then generates 3D chromatin ensembles using identified specific interactions as spatial constraints. Application to $\alpha$-globin locus shows that these constraints (~20%) drive the formation of ~99% all experimentally captured interactions, in which ~30% additional to the imposed constraints is found to be specific. Many novel specific spatial contacts not captured by experiments are also predicted. A subset, of which independent ChIA-PET data are available, is validated to be RNAPII-, CTCF-, and RAD21-mediated. Their positioning in the architectural context of imposed specific interactions from nC-SAC is highly important. Our results also suggest the presence of a many-body structural unit involving $\alpha$-globin gene, its enhancers, and POL3RK gene for regulating the expression of $\alpha$-globin in silent cells.**

## INTRODUCTION

A central problem in biology is understanding the spatial organization of the genome inside a cell nucleus and how 3D genome folding dictates important cell activities such as gene expression (1). Recent development of chromosome conformation capture (3C) and related techniques (4C, 5C, Hi-C) enabled large-scale discovery of long-range chromatin looping interactions among distant chromosomal elements (2–7). The discovery of Topologically Associated Domains (TADs) with elevated chromatin interactions (8,9) suggests a detailed structural network involving binding of architectural proteins (10). These findings point to likely 3D structural units of chromatin that accommodate spatial clustering of different regulatory elements and transcription factors important for cell activities.

Chromatin is highly dynamic and experiences significant conformational changes (11). As 3C data are averaged over cell populations and may reflect a mixture of different conformations at a particular moment, it is important to uncover an ensemble of 3D structures of a gene locus and assess the occurrence of different structures that collectively best describe the bulk measurements (12). This would enable precise structural measurements and identification of spatial organizational units of genomic elements.

To overcome the limitations of the pairwise nature of chromosome conformation capture data and to gain detailed mechanistic understanding of gene regulation, there have been significant efforts in constructing 3D structures of chromatin. Current 3D polymer models can be categorized as either thermodynamics-driven or data-driven models (13). Thermodynamics-driven models are based on minimal physical assumptions and sometimes incorporate empirically additional information such as CTCF (14–16), epigenetics states (14,17), or 3C related data (18,19). They have revealed important information on the general folding principles of genome (5,14–26). For example, a recent model based on energy landscape can reproduce experimentally captured interaction patterns of human genome using parameters derived from one chromosome and transferred to other chromosomes (14). The formation of TADs (16,17) can also be predicted. However, such models are often general models that do not provide detailed spatial structures necessary for understanding the underlying mechanism of differential gene expression (5,16,17,20–26), or provide details only on CTCF interactions (15). Others have limited

[*]To whom correspondence should be addressed. Tel: +1 312 355 1789; Fax: +1 312 996 5921; Email: jliang@uic.edu
[†]These authors contributed equally to this work.
Present address: Gamze Gürsoy, Yale University, Molecular Biophysics, and Biochemistry, Computational Biology and Bioinformatics.

resolution for capturing finer structural differences of a small locus (14,18). While there has been important recent success in identifying specific driver interactions for small loci (∼150 kB) (19), it requires significant amount of simulation that is difficult to scale up.

Data-driven models generate 3D ensemble of chromatin chains using 3C-related (4C/5C/Hi-C) data and can provide rich information on biological properties of genomic elements (27–36). For example, experimentally obtained interaction patterns can be reproduced computationally with simple assumptions (27,28,31–33,36), with co-localization of co-expressed genes uncovered (33) and TAD boundaries predicted (36). However, it is challenging to computationally generate well-sampled ensembles of high-resolution chromatin chains and to quantify the occurrence of different chain configurations (29,30). In addition, the majority of these models have been designed to study overall genome organization and cannot provide detailed information on a specific locus (27,28,31–33,36). Furthermore, current constraint-based models cannot distinguish specific interactions from non-specific interactions.

In this study, we describe the *n* Constrained-Self-Avoiding Chromatin (nC-SAC) method for constructing 3D ensemble of chromatin structures from 3C-related measurements and for identifying specific interactions that may be important to drive the formation of spatial structures of gene loci. We focus on the α-globin locus and predict detailed configurations of ensembles of chromatin chains based on 5C-derived constraints (29).

As measurements from 3C-related techniques impose a large number of proximity constraints, existing data-driven methods based on spatial constraints have difficulties in distinguishing specific interactions from non-specific interactions (13), which arise from random interactions due to non-specific collision of chromosomal regions to one another (37). We developed a two-staged approach to overcome this problem. In the first stage, we examine chromatin interactions arising solely from the effects of the finite volume of cell nucleus and the self-avoiding nature of the chromatin. Specifically, we generate 100,000 random self-avoiding chromatin chain conformations inside the crowded cell nucleus, which are used as the null model. The nC-SAC model then identifies the most significant 5C interactions and separate them from non-specific interactions in the α-globin locus. In the second stage, the significant interactions are imposed as spatial constraints for construction of 3D models of chromatin chains. By using the geometrical Sequential Importance Sampling (g-SIS) technique, we overcome severe sampling problems and generate large ensembles of 3D chromatin chains of the α-globin locus for two cell lines with different expression levels. These ensembles satisfy ∼90% of the imposed constraints of significant 5C interactions.

We further examine the functional consequences of the structures of α-globin locus in gene regulation. Our structural model predicts a large number of novel looping interactions with spatial details that were not captured by the original 5C experiment due to lack of primer coverage. A subset of our predicted interactions are shown in two independent ChIA-PET studies to be mediated by proteins such as CTCF, RNAPII, RAD21 and are associated with concurrent histone modifications (38,39). Analysis of a second null model, where the position of interaction pairs are randomized, indicates that the association of these newly discovered interactions with functional elements is highly significant. Our model further suggests the existence of a many-body structural unit involving α-globin gene, enhancers HS40/46/48, and POL3RK gene for regulating α-globin expression in the silent cell line. In addition, our models uncover global differences in the spatial structures of the α-globin in cells with high and low expression. More particularly, our findings suggest that a homogeneous and dominant structural population of the locus may be associated with the high expression level of the α-globin. Furthermore, interactions identified by the nC-SAC method, including predicted novel interactions, may play important roles as driver interactions for the formation of the ensemble structures of α-globin.

## MATERIALS AND METHODS

The overall computational pipeline of nC-SAC is illustrated in Figure 1. In stage 1, the interaction frequencies obtained from the 5C study (29) are compared to the interaction frequencies of a random C-SAC ensemble (Figure 1A and B) (40). For this purpose, an ensemble of 100,000 C-SAC chains (40) confined to a sphere is generated (Figure 1A). This ensemble is then bootstrapped 1000 times and the *P*-value of observing the experimentally captured interaction frequency in the random ensemble is calculated. The obtained *P*-values are then corrected for multiple hypothesis testing (Figure 1C). A direct relationship between interaction frequency and spatial distance was observed experimentally (41). Thus, significant 5C interaction frequencies at the False Discovery Rate of $\alpha < 5\%$ (Figure 1D) are converted into spatial distances using a half-Gaussian model (Figure 1E). In stage 2, an ensemble of 10,000 3D chromatin chains that satisfy these spatial distance constraints are then generated using the technique of g-SIS (Figure 1F). A full resolution interaction map is computed from the generated ensemble of structures (Figure 1G).

### Mapping 5C data on to a polymer chain

We model the α-globin chromatin chain as a polymer chain under confinement, following our previous study on the volume effects of cell nucleus on human chromosome folding (40). In this model, we represent the chromatin as a collection of beads and the chain is constrained by 5C interaction frequencies and the confinement of the cell nucleus. We divide the 500 kb locus into 184 beads, each corresponds to 2.7 kb DNA. We used fragment units to mimic the HindIII restriction fragmentation. If a HindIII fragment is shorter than five beads (∼13.5 kb), it is represented by one fragment unit. The HindIII fragments longer than five beads are divided into multiple fragment units. Each fragment unit is modeled as a rigid body with a maximum of five beads. We call the last bead in each unit a node, and the node number is used as the identifier of the unit (Figure 2A). In total, we obtain 54 rigid fragments with different lengths, where the maximum length of the unit is kept at five beads to conform with the chromatin bending property, which corresponds to a persistence length of 150 nm. An alternative

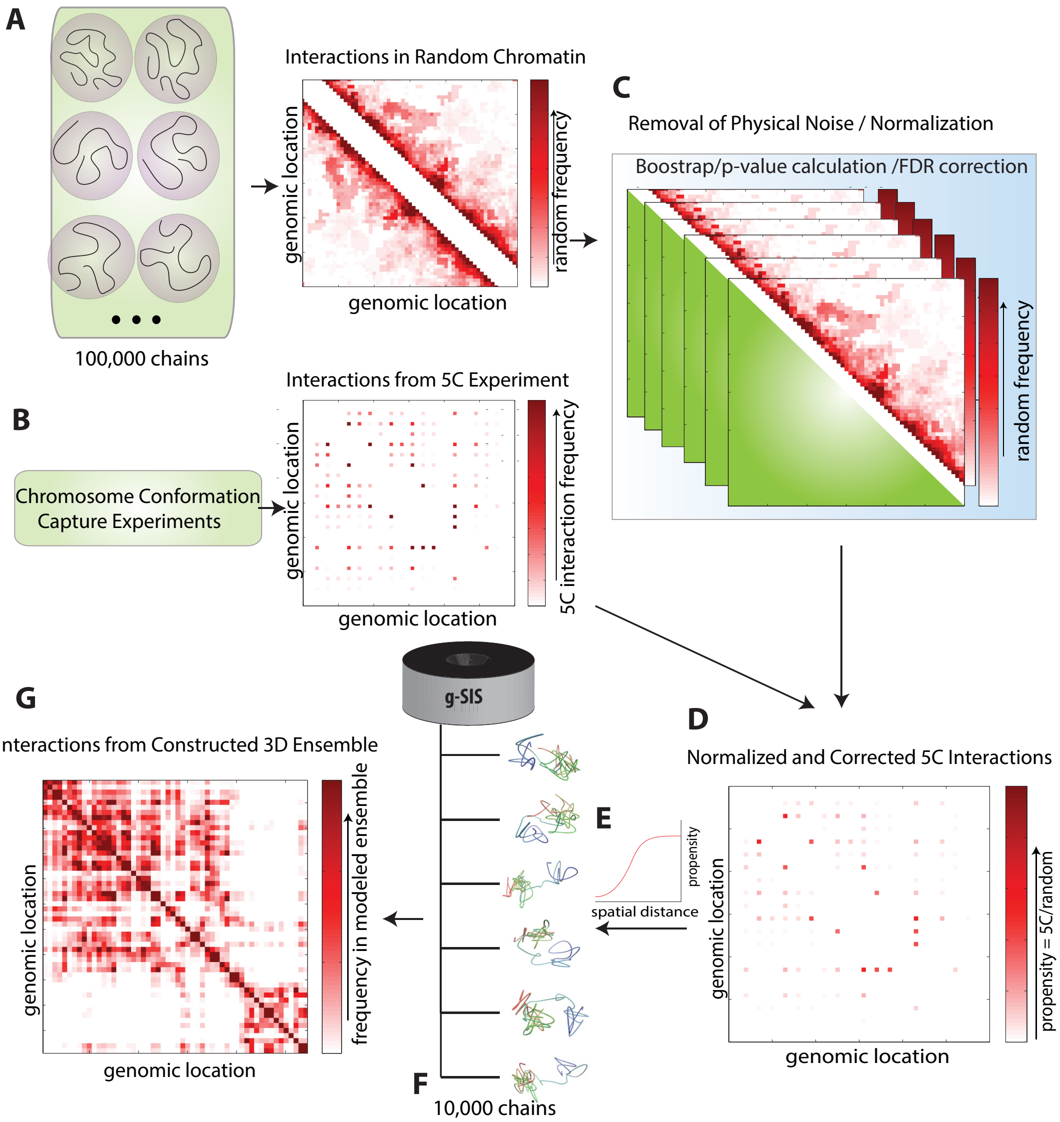
A
100,000 chains
Interactions in Random Chromatin
genomic location
genomic location
random frequency
C
Removal of Physical Noise / Normalization
Boostrap/p-value calculation /FDR correction
random frequency
B
Chromosome Conformation Capture Experiments
Interactions from 5C Experiment
genomic location
genomic location
5C interaction frequency
g-SIS
G
Interactions from Constructed 3D Ensemble
genomic location
genomic location
frequency in modeled ensemble
D
Normalized and Corrected 5C Interactions
E
propensity
spatial distance
genomic location
genomic location
propensity = 5C/random
F
10,000 chains

**Figure 1.** The nC-SAC computational pipeline to predict structural ensembles of chromatin chains from 5C data. (**A**) The interactions of a random 3D ensemble of $10^5$ C-SAC chains in cell nucleus generated to obtain a contact matrix of random interactions. (**B**) 5C interactions are compared with (**C**) 1000 bootstrapped random contact matrices to calculate the *P*-value of each interaction. (**D**) After FDR adjustment for multiple hypothesis testing, only significant 3D interactions are considered. (**E**) These remaining significant 5C interactions are normalized and converted into distances using a half-Gaussian model. (**F**) An ensemble of $>10^4$ 3D-chains of the locus is then generated under these constraints and (**G**) a full resolution contact map is computed and compared to (B).

approach to model persistence length is that of (27), where an explicit bending energy is introduced so that a distribution of bending angles that conform to known persistence length of chromatin is obtained.

### Stage 1: Identification of specific physical interactions

We first discard 5C interactions associated with short (<2.7 kb) fragments as they are considered to be unreliable (42). We also discard interactions between consecutive fragments as they are likely due to proximity effects (2). An ensemble of 100,000 randomly folded polymer chains that have the same physical properties as α-globin locus (53 nodes, 183 monomers) are generated inside a confined space of nucleus using previously described C-SAC model (40) to asses the statistical significance of each 5C interaction. The size of the confined space is that the volume of 500 kb length DNA would occupy, which is calculated to be proportional to a diploid human nucleus. Following experimentally determined threshold (19), any two nodes with distance less than 80 nm is considered to be in contact. After normalization and calculation of propensity for each interaction (see Supplementary Methods), the *p*-value for each 5C interaction is calculated by bootstrapping the random ensemble 1000 times. After a correction of multiple hypothesis testing through the FDR (43) analysis with an $\alpha = 0.05$, any interaction that passes the FDR is considered to be a specific interaction. We use a simple Gaussian function to con-

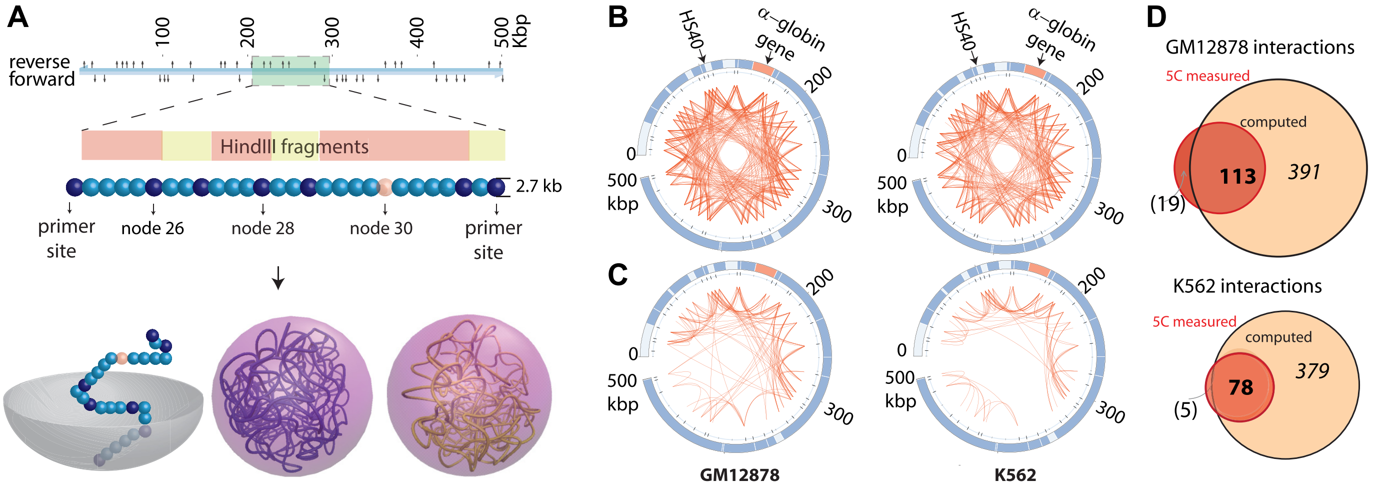

**Figure 2.** Mapping 5C interactions onto nC-SAC model chromatin chains, identifying non-specific interactions, and predicting novel interactions between genomic elements in α-globin locus. (**A**) Mapping the 500 kb α-globin locus and 5C interactions onto the C-SAC polymer model of chromatin chain. Up and down arrows in the linear diagram of the α-globin locus represent reverse and forward primers at ends of HindIII fragments (29), respectively. Fragments between primer sites 25 and 32 are enlarged to demonstrate details of the C-SAC model. Alternating fragments are shown in pink and yellow. HindIII fragments are mapped onto fragment units, with which bending can occur at the primer sites (darker blue) or every 6-th bead (pink) for fragments mapped onto one or more units. Ensemble of random C-SAC chromatin chains are generated through chain-growth one bead at a time in a confined sphere representing the cell nucleus (40). Representative partial and full C-SAC chains in spherical confinement are shown. (**B**) All reported 5C interactions between elements in the α-globin locus (29). (**C**) Remaining significant 5C interactions after non-specific interactions are excluded. (**D**) Comparison of statistically significant 5C interactions (red circle) and interactions predicted by the nC-SAC model (beige circles). The numbers of 5C interactions captured by nC-SAC are in bold, and those not captured are in parenthesis. Among predicted interactions (beige circles), many are novel predictions (in italic) that are not measured in the original 5C study.

vert propensity of each significant interaction into a spatial distance between nodes (Supplementary Methods).

### Stage 2: Significant 5C interactions as physical constraints

The nC-SAC model is an extension of the C-SAC model, where $n$ spatial distances ($d^{5C}_{cell}(i, j)$) are used additionally as constraints in the geometrical Sequential Importance Sampling procedure of the original C-SAC method during the chain growth process (40,44–48). The configuration x of a full chromatin chain with $N$ nodes, with the location of the $i$-th node denoted as $x_i = (a_i, b_i, c_i) \in \mathbb{R}^3$, is: $\mathrm{x} = (x_1, \cdots, x_N)$. Our goal is to generate chromatin conformations that maximize the target distribution $\pi(\mathrm{x})$, which reflects the extent constraints identified in Stage I are satisfied. Our target distribution $\pi(\mathrm{x})$ is the distribution of chromatin chains, in which the spatial distances $d^{pred}_{cell}(i, j)$ between nodes $i$ and $j$ are equal to the spatial distances derived from significant 5C interaction frequencies ( $d^{5C}_{cell}(i, j)$), while ensuring the self-avoiding property. To generate a chromatin chain, we grow the chain one node at a time, by using a $k$ = 640-state off-lattice discrete model (40,44–48). The new node added to a growing chain with the current node located at $x_t$ is placed at $x_{t+1}$, which is a fragment unit $L_p$ distance away from $x_t$. $x_{t+1}$ is taken from one of the unoccupied $k$-sites neighboring $x_t$ according to a probability distribution that favors the $d^{pred}_{cell}(i, j) = d^{5C}_{cell}(i, j)$ and the self-avoiding property, namely, $x_i \neq x_j$ for all $i \neq j$. As satisfying the $\{d^{5C}_{cell}(i, j)\}$ constraints where $(i, j)$ pairs are far away in genomic distance is extremely challenging, we introduce a look-ahead biasing strategy to select the $(i, j)$ pair from available empty neighboring sites that do not have constraints. We keep track of this bias and assign each successfully generated chain a proper weight $w(\mathrm{x})$, which is calculated as the deviation of sampling distribution with respect to the target distribution $\pi(\mathrm{x})$ (40,49–50). While bias is introduced to increase the sampling efficiency, it generates chains that diverge from the target distribution. Therefore, the weight of a successfully generated chain is used for correcting the bias introduced. This weight is calculated using Supplementary Equation (8). More details of biasing by look-ahead can be found in (49,50).

## RESULTS

### Identifying specific 5C interactions

During the construction of 3C chromatin libraries, formaldehyde treatment can covalently link genomic elements within certain spatial distances, regardless whether specific interactions exist (37). A significant number of such interactions arise from the self-avoiding nature of chromatin chains confined in the crowded cell nucleus (40,22). As loci from different chromosomes can coexist inside the cell nucleus, the nuclear confinement as well as the excluded volume effects result in limited available space for individual locus. We hypothesized that many 5C interactions are of non-specific nature, while others are specific in the sense that they correspond to certain architectural constraints. In a first stage, we identify such specific and non-specific interactions. We generated an ensemble of 100,000 random C-SAC chains (40) (Figure 2A). Without *a prior* information, we assume the 500 kb ($5 \times 10^5$ bp) α-globin locus occupies $(5 \times 10^5)/(6 \times 10^9)$ of the available space of $\sim 7.5^3$ μm$^3$ inside an average cell nucleus of ∼$10^3$ μ$m^3$ with a diploid human genome size of $2 \times 3 \times 10^9$ bp. This corresponds to a sphere of a diameter of 330 nm.

We bootstrap the chains in the random ensemble to generate 1000 ensembles of 100,000 C-SAC chains and calculate the probabilities of observing the same normalized 5C interaction frequencies in these random ensembles and used these probabilities as *P*-values, which are then subject to correction of multiple hypothesis testing at the False Discovery Rate (FDR) of $\alpha < 5\%$ (see Methods and Supplementary for more details). A total of 293 of 425 experimentally captured 5C-interactions (77%) in the GM12878 cell line and 284 of 367 5C-interactions (87%) in the K562 cell line are not statistically significant and are therefore considered to be non-specific and not used as spatial constraints (Figure 2B and C). Only 23% and 13% of the 5C interactions are found to be significant in the GM12878 and K562 cell lines, respectively. Recognizing that the available space may not be perfectly spherical, we use a stringent FDR criteria to ensure that we only identify the most significant interactions, which will be present even if there are deviations from the ideal spherical shape. Our explorations showed that imposing more stringent criteria of $\alpha = 0.01$ would result in excluding only one more interaction for each of the cell lines.

We then asked whether the 5C interactions identified as specific are generally consistent with the overall observation of long-range chromatin interactions of the locus. Interactions of α-globin gene with its enhancer HS40, as well as interactions with neighboring hypersensitive sites HS48 and HS46 were identified as key factors determining the expression levels of α-globin gene in globin expressing mouse cells in prior knock-out and 3C studies (51,52). We found that pairwise interactions between the α-globin gene and the regulatory elements HS40, HS46 and HS48 are among the specific interactions (Supplementary Figure S1). Moreover, ChIP-Seq experiments (53) reveal that these nodes contain RNAPII peaks (Supplementary Figure S5). Additional ChiA-PET experiments (38,39) reveal that those interactions are likely mediated by RNAPII (Supplementary Table S2, between node pairs 11–22 and 12–22), CTCF (Supplementary Table S4, between node pairs 11–22, 12–22 and 13–22) as well as RAD21 (cohesin) (Supplementary Table S5, between node pairs 11–21 and 12–21).

## A few specific interactions drive the formation of majority of 5C contacts

In Stage 2, 3D structural models of the α-globin locus that satisfy 5C interactions identified as specific are generated using nC-SAC method. Following previous studies (6,29,30,33) as well as experimental observations (41), we assume an inverse relationship between 5C frequencies and spatial distances, and employ a simple half-Gaussian model to map frequencies of significant 5C interactions to spatial distances between nodes (detailed in Materials and Methods and SI). These spatial distances are then regarded as physical constraints that the 3D chromatin chains need to satisfy. Two separate ensembles of 10,000 chromatin chains of the α-globin locus are then generated for the GM12878 and the K562 cell lines (Supplementary Figure S2).

The α-globin chromatin ensembles generated using the specific 5C interactions miss only two interactions out of 425 5C interactions in the GM12878 cell line and miss only two interactions out of 367 5C interactions in the K562 cell line. That is, 99% of all 5C interactions are captured in the constructed 3D ensembles. To further analyze the biologically specific interactions in the predicted 3D ensemble, we discarded any interaction that are present in the random C-SAC ensemble with high probability by using the same bootstrapping technique of Stage I at the FDR level of 5%.

We assessed how well the specific interactions of generated 3D ensembles of chromatin chains satisfy the constraints imposed, we found that they capture 78 (94 %) and 113 (86 %) of the imposed significant 5C interactions for K562 and GM12878 cell lines, respectively (Figure 2D). Further investigation showed that 74 and 122 additional interactions not imposed as constraints are also re-captured as a consequence of imposed constraints. When all 5C interactions are considered, nC-SAC identifies 41% and 55% of them as specific for K562 and GM12878 cell lines, despite that only a small fraction (13–23%) of them are used as constraints. Here captured interactions in the nC-SAC ensemble are interactions between nodes identified to have distances below the 80 nm threshold and with significantly higher propensity compared to random ensemble at the level of 5% FDR rate. These observations suggest that while our conservative approach for excluding non-significant interactions are stringent and only the most significant interactions are imposed as constraints, 3D ensembles of chromatin chains generated by nC-SAC can uncover many additional 5C interactions that are physical and may be biologically relevant.

These results also shows that 13–23% of the raw 5C interactions are adequate to give rise to additional 30% of the 5C interactions. That is, formation of a number of experimentally captured interactions is driven by a small number of specific interactions. Furthermore, the predicted chromatin chains of the α-globin locus exhibit many novel interactions not present in the original 5C data (278 and 301 interactions in the GM12878 and K562 cell lines, respectively).

*nC-SAC uncovers structural differences of* ***α****-globin locus.* There are global structural differences in the organization of the α-globin gene locus between K562 and GM12878 cells, as seen in the heatmaps of spatial interactions from predicted α-globin chains (Figure 3A). Overall, α-globin locus of the silent GM12878 cell line forms a single compact chromatin globule. In contrast, chains of the active K562 cell line are extended, forming two non-interacting globules, which exhibit two separate domains in the heatmap (Figure 3A). These findings are consistent with previous results (29). Additional global and local structural differences have been captured using a density based clustering algorithm (54) (Supplementary Methods). Two types of partitioning of the ensemble of 3D chromatin chains of each cell line into clusters based on (i) the radius of gyration of the model chromatin chains and (ii) the root mean square deviations derived from corresponding pairwise distances between two within-chain nodes are performed to provide information on global heterogeneity of chromatin chains based on their openness or compactness as well as how well chromatin

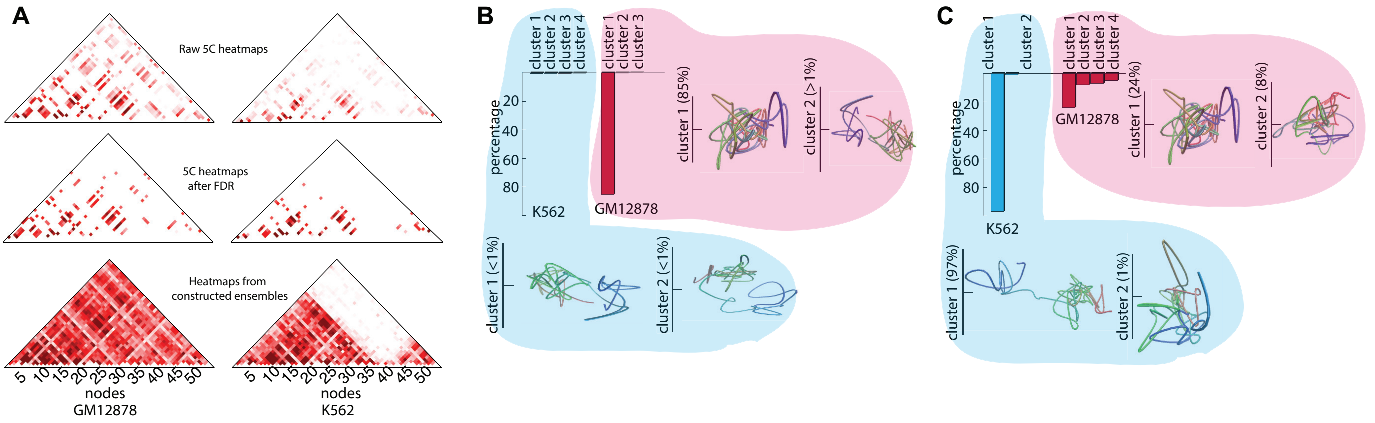

**Figure 3.** Ensembles of predicted 3D chromatin chains of the α-globin locus. Interactions between genomic elements of the α-globin locus from predicted structural ensembles of 10 000 chromatin chains in the silent GM12878 and the active K562 cell lines. (**A**) Heatmaps of spatial interactions of α-globin locus including raw 5C counts, most significant 5C counts after exclusion of non-specific interactions upon FDR correction, and interactions from the modeled structural ensembles. The normalized frequency of *i–j* interactions is color coded. Red intensity indicates increased frequency. (**B**) The histogram (top) shows the proportion of structures associated with different structural clusters (K562, blue; GM12878, red), when clustered by the differences in radius of gyration between the chains. The predominant three dimensional structures associated with structural clusters 1 and 2 are also shown for both cell lines. (**C**) The histogram (top) shows the proportion of structures associated with different structural clusters (K562, blue; GM12878, red), when clustered by the pairwise differences between nodes of modeled chromatin chains. The predominant three dimensional structures associated with structural clusters 1 and 2 are also shown for both cell lines.

chains in the ensemble maintain detailed 5C interactions that may be necessary for gene expression.

The global clustering results showed that the ensemble of the non-expressing GM12878 cell line is homogeneous in radius of gyration with overall a small number of clusters (a total of ∼26), with the most populated cluster accounting for ∼85% of the chromatin chains in the ensemble. In contrast, the ensemble of the α-globin expressing K562 cell line is globally diverse and contain many clusters (a total of ∼299) with most prominent cluster accounting for only <1% of the whole ensemble (Figure 3B). When the second type of clustering is performed, the ensemble of the α-globin expressing K562 cell line is remarkably homogeneous with overall a small number of clusters (a total of ∼13), with the most populated cluster accounting for ∼97% of the chromatin chains in the ensemble. In contrast, the ensemble of the non-expressing GM12878 cell line is diverse in detailed local interactions, with many different clusters (a total of ∼148), with the most prominent cluster accounts for only ∼24% of the whole ensemble (Figure 3C). These results emphasize the compactness of the 3D chains in the non-expressing cell line and the openness of the 3D chains in the expressing cell lines, consistent with previous results (29). Our results also indicate that there are significant differences in subpopulation heterogeneity of chromatin chains in local interactions between the two cell lines.

*nC-SAC predicts novel interactions mediated by CTCF, RNAPII and RAD21 proteins.* In the context of the constraints originating from the specific frequencies, we predict the existence of additionally 504 and 457 long-range interactions in the α-globin locus for GM12878 and K562 cell lines, respectively. These specific interactions are observed significantly more often in the nC-SAC ensemble compared to the random C-SAC ensemble at the FDR level of 5%. To determine the biological relevance of these interactions, which are not imposed as constraints, we examined results from two independent ChIA-PET studies of K562 cells (38,39) and a ChIA-PET study of GM12878 cells (39). These ChIA-PET studies revealed looping interactions in the α-globin locus mediated through RNAPII, CTCF and RAD21 binding, as well as interactions associated with histone modifications (38,39).

Among the 68 RNAPII-mediated interactions in K562 cells detected by ChIA-PET (38) (Figure 4A, blue circle in the Venn diagram and Figure 4B1, blue and gray arcs), 33 are also predicted by nC-SAC (Figure 4A, orange circle). Notably, 21 of the 33 predicted interactions are novel interactions absent in the 5C measurements (Figure 4B2, red arcs) and 12 are interactions captured by 5C measurements (Figure 4B3, green arcs). Among the 35 RNAPII-mediated interactions undetected by nC-SAC (Figure 4B1, gray arcs), 26 have no primer coverage and therefore are not reflected in the 5C data. The remaining 9 RNAPII-mediated interactions have low or no 5C interactions, imposing very weak constraints for our model. A separate ChIA-PET study (39) revealed two more RNAPII mediated interactions between α-globin gene (node 21) and HS40 (node 12), which are also predicted by nC-SAC and are present in 5C measurements (Supplementary Table S1).

We also examined CTCF-mediated interactions in K562 cells detected by ChIA-PET (38). Among the 11 reported interactions (Figure 4C, blue circle in the Venn diagram and Figure 4D1, blue and gray arcs) (38), eight are predicted by the nC-SAC model (Figure 4C, orange circle). Of those, six are absent in the 5C measurements (Figure 4D2, red arcs) and two interactions are captured in 5C measurements (Figure 4D3, green arcs). The three CTCF-mediated interactions detected by ChiA-PET but undetected by nC-SAC

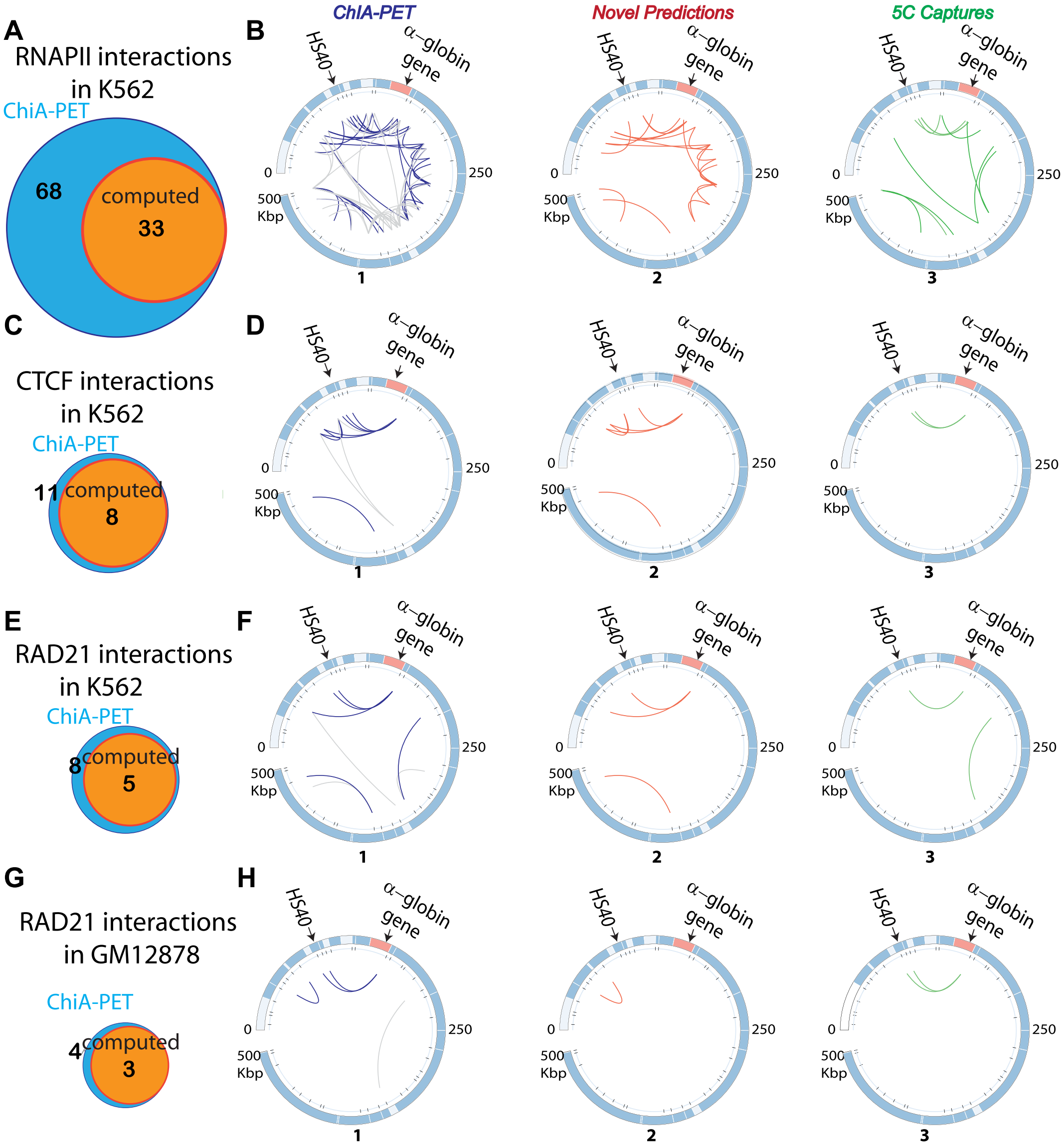

**Figure 4.** Predicting novel interactions between genomic elements in α-globin locus and validation of their biological relevance. (**A**, **C**, **E**, **G**) Comparing looping interactions detected by ChIA-PET (38) and nC-SAC 3D ensemble predicted interactions in K562 cells. The Venn diagrams show ChIA-PET measured (blue circles) and nC-SAC predicted (orange circles) interactions. (**B1**, **D1**, **F1**, **H1**) The circos diagrams show interactions detected by ChIA-PET (blue arcs for captured interactions by 3D model and gray arcs for interaction that are absent in the 3D model), (**B2**, **D2**, **F2**, **H2**) nC-SAC predicted interactions detected by ChIA-PET but absent in 5C (red arcs), (**B3**, **D3**, **F3**, **H3**) interactions predicted by nC-SAC and captured by the 5C and ChIA-PET techniques (green arcs).

(Figure 4D1, gray arcs) either have no 5C frequency or have no primer coverage, hence impose no constraints for our model.

In addition, we examined RAD21-mediated interactions in K562 cells detected by a recent ChIA-PET study (39). Among the eight reported interactions (Figure 4E, blue circle in the Venn diagram and Figure 4F1, blue and gray arcs), five are predicted by the nC-SAC model (Figure 4E, orange circle), three of them are novel interactions that are absent in 5C measurements (Figure 4F2, red arcs) and two interactions are captured in 5C measurements (Figure 4F3, green arcs). The three RAD21-mediated interactions detected by ChiA-PET but undetected by nC-SAC (Figure 4F1, gray arcs) have no 5C coverage, imposing no constraints for our model.

We further examined RAD21-mediated interactions in the silent GM12878 cells detected by a ChIA-PET study (39). Among the four reported interactions (Figure 4G, blue circle in the Venn diagram and Figure 4H1, blue and gray arcs), three are predicted by the nC-SAC model (Figure 4G, orange circle), including one novel interaction absent in the 5C study (Figure 4H2, red arcs), as well as two interactions captured by 5C (Figure 4H3, green arcs). The only undetected interaction (Figure 4H1, gray arcs) has no 5C coverage, imposing no constraints for our model.

Overall, our nC-SAC method has predicted 52% of the 68 RNAPII-mediated interactions, 75% of the 11 CTCF-mediated interactions, 62% of the 8 RAD21-mediated interactions in K562 cell line, 86% of the 7 interactions that are associated with histone modifications in K562 cell line, and 80% of the 5 RAD21-mediated interactions in GM12878 cell line (Supplementary Tables S1–S6). In total, 89 interactions are detected by ChiA-PET in K562 cell line and 52 of them are among 457 predicted significant interactions. To assess the significance of the functional enrichment of predicted interactions, we constructed a null model, where the position of the functional pairs are randomly redistributed using the same 3D ensemble of chromatins chains constructed under the same 5C constraints. Among the 100,000 instances of random redistributions of the 89 ChIA-PET detected interactions, the *P*-value of 52 or more interactions are identified as significant by nC-SAC is small ($P < 10^{-5}$, Supplementary Figure S3). These results suggest that our predicted interactions are indeed functional interactions and are specifically located with respect to the 5C constraints.

*nC-SAC predicts novel interactions associated with concurrent histone modifications.* We then examined the interactions that are found to be associated with histone modifications in K562 cells in a recent ChIA-PET study (39). Among the seven reported interactions, six of them are predicted by nC-SAC model, three of them are novel interactions that are absent in 5C measurement, and three of them are captured by the 5C study. The only undetected interactions has no 5C coverage (Supplementary Table S1).

Our results are consistent with the finding that the positioning of functional genomic elements in chromatin loops are important for the regulation of gene expression (55). Our results are also consistent with the importance of specific positioning of CTCF with respect to functional genomic elements (15). Furthermore, our results suggest that some of the small number of interactions used as input constraints in constructing nC-SAC models may function as driver interactions, whose introduction result in the formation of specific interactions among functional sites.

*nC-SAC predicts detailed 3D structural interactions.* Expression of the α-globin gene is thought to be regulated through enhancer-promoter interactions (29). In accord, the interaction between the α-globin gene and enhancers HS40/46/48 are found in 90% of predicted chains of the active K562 cells. This is expected, as the interaction between the α-globin gene and enhancers HS40/46/48 are identified in Stage I and imposed in Stage II. However, this represents an increase by a factor of only 1.29 compared to the silent GM12878 cells, as this interaction is also present in 69.8% of predicted chains of the GM12878 cells (Figure 5A). Our finding is consistent with a previous ChIA-PET study, in which interactions between HS40 and α-globin gene is found to be mediated by RAD21 in the silent GM12878 cell line (Figure 4G and H) (39). These observations indicate that α-globin promoter-enhancer interactions alone do not determine the expression level and additional regulatory elements may be at play.

We examined nC-SAC predicted 3D structures for K562 and GM12878 cells to assess the presence of other looping interactions, which may regulate α-globin expression (Figure 5A–D). While it is difficult to compare absolute interaction frequencies between cell lines, we can compare the relative fractions of chromatin chains containing specific interactions in each cell line.

We first identify spatial interactions that both α-globin gene and enhancers participate concurrently in the two cell lines (Supplementary Figure S4). This allowed us to further examine higher-order interactions involving other nodes beyond α-globin and enhancers, but occurring exclusively in one cell line only. We overlapped epigenetic profiles of each of these differentially interacting nodes and identified those that are associated with epigenetic marks (Supplementary Figures S4 and S5). From this analysis, our nC-SAC study predicts that the POL3RK gene engages in a three-way interaction with the α-globin gene and enhancers in 70% of α-globin chromatin structures from GM12878 cells. In contrast, the POL3RK gene has a much lower three-way interaction frequency (18%) with the α-globin gene and enhancers in K562 cells (Figure 5A). The predicted interaction between POL3RK and enhancers was not detected in the original 5C study due to primer design strategy. (29). With explicitly generated 3D structures, we can measure the exact Euclidean distances between genomic elements in individual chains and can calculate their ensemble averages. We found chains from GM12878 cells with POL3RK:α-globin:enhancers three-body interaction all have average pairwise distances between elements (50.1 ± 20 nm, 62.4 ± 18 nm, and 80.0 ± 5nm) shorter or near the threshold of interaction (∼80 ± 5 nm) given in previous studies (19) (Supplementary Table S7, Figure 5D). In contrast, the averaged spatial distances of POL3RK:α-globin (∼135 ± 20 nm) and POL3RK:enhancers (∼140 ± 18 nm) in active K562 cells are both much longer than this threshold (Supplementary Table S7).

We speculate that the three-way looping interaction of POL3RK with the α-globin gene and enhancers may occlude access of transcription factors to the α-globin transcriptional elements, thus silencing the α-globin expression (Figure 5B–D). This denial of access could be aggravated when transcription factors bound to the POL3RK gene occupies much of the available space. This scenario is consistent with epigenetic data, in which POLR3K in the silent GM12878 cells is enriched for transcription factors binding Pu.1 and Sp1 and for histone modifications H2A.Z and H3Kme2, both of which are related to transcriptional activation (Supplementary Figure 5) (53). Furthermore, it is also consistent with the observed lack of H3K4me2 modifications on α-globin enhancers in the silent cells, which is related to abundance of transcription factor binding, as well as with the lack of RNAPII enrichment, which is related to absence of gene expression (Supplementary Figure S5).

## DISCUSSION

We describe the nC-SAC method that can transform 2D maps of Chromosome Conformation Capture frequencies of interactions into a population of 3D chromatin chains. Our method identifies the most significant spatial interactions, overcomes the sampling problem, and generates a large number of properly sampled self-avoiding chromatin chains that satisfy constraints imposed by 5C interactions.

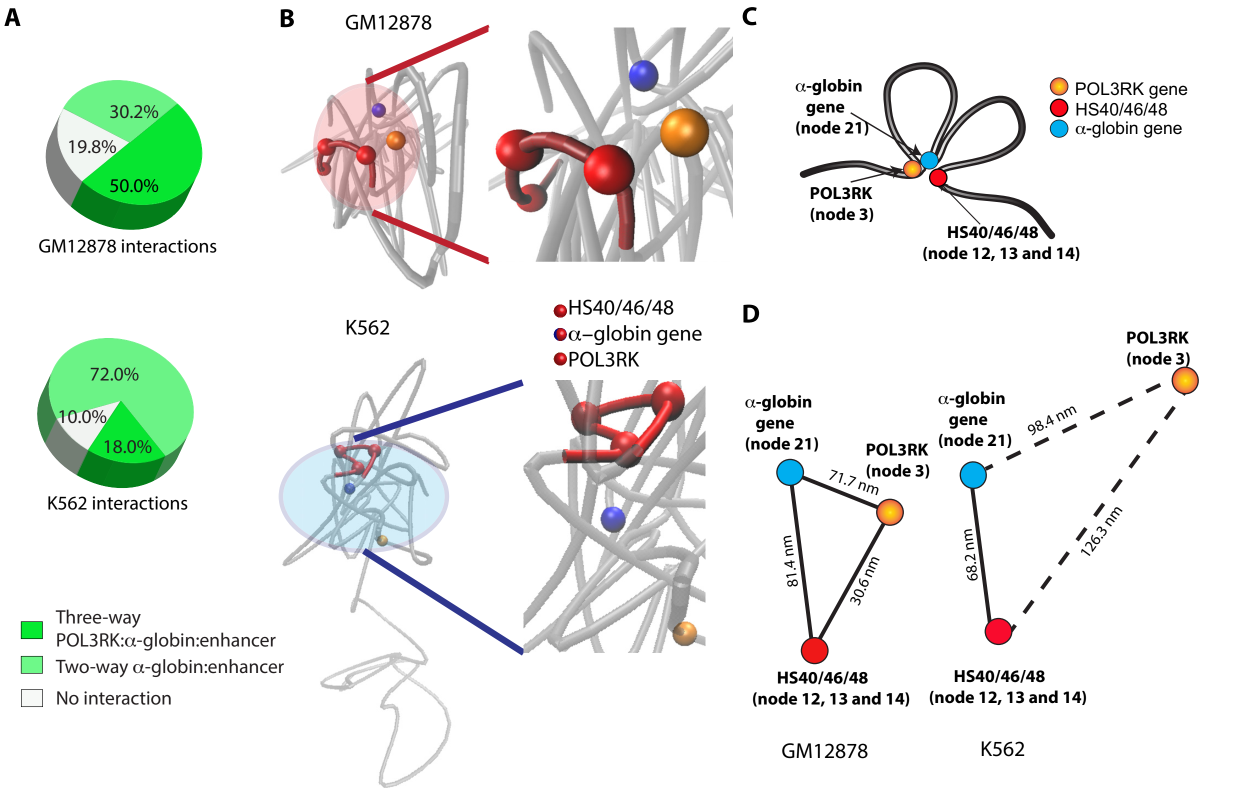

**Figure 5.** Three-way interaction of POL3RK:α-globin gene:enhancers is likely a unique feature in the non-expressing GM12878 cells. (**A**) Pie charts depicting the percentages of the ensembles that have two way (α-globin:enhancer, in light green) and three-way (POL3RK:α-globin gene:enhancers, in dark green) interactions in both GM12878 and K562 cell lines. (**B**) The spatial structures of α-globin locus chromatin were reconstructed from nC-SAC predicted 3D chromatin chains, with the enhancers HS40/46/48 (red), POL3RK (orange), and the α-globin gene (blue) depicted. The structures shown are drawn from the most populated clusters of GM12878 and K562 cells. (**C**) A schematic representation of the three-body interaction of α-globin gene (blue), enhancer (red), POL3RK (orange) observed in GM12878 cells. (**D**) The spatial distances between the enhancers HS40/46/48 (red), POL3RK (orange), and the α-globin gene (blue) of the three-way interaction unit in the representative structure depicted in (B).

Although its resolution is limited to that of the HindIII fragments in this study (5–13 kb) and no direct information is provided on chromatin dynamics, this method enables us to examine structural properties of the α-globin locus, allowing structural and distance measurements at the population level in a manner consistent with the basic requirement of the physical chromatin chains and the 5C interactions.

While the structures of α-globin locus based on 5C interactions were modeled in the original 5C paper (29), there were several questions remained unanswered in this early study. First, the expected interactions were determined by the method of LOESS smoothing (29), which requires availability of 5C interaction frequencies at each interval of genomic distance. Due to the sparsity of the HindIII fragments, obtaining an adequate sample size for accurate LOESS smoothing is challenging. This affects the accuracy of structural modeling. For example, Bau et. al predicted that the interactions between HS40 and α-globin gene in GM12878 cell line is as would be expected by the LOESS smoothing. However, a recent ChiA-PET data found the interaction between HS40 and α-globin locus as statistically significant and are mediated through RAD21 (39). Using our stringent FDR criteria, this interaction in GM12878 cell line was found to be significant by nC-SAC, in agreement with the ChiA-PET observations. Second, each HindIII fragment in ref. (29) was modeled as a single sphere whose size is proportional to its genomic length. This resulted in enormous beads with unrealistic excluded-volumes at places in the polymer model. For example, a HindIII fragment that contains 30kb DNA would correspond to a bead ∼300 nm according to (29). It also prevented the prediction of detailed interactions beyond the resolution of the 5C data. Third, HS40 and α-globin genes were closely interacting in the locus in K562 cells according to the (29), but with a distance of 159.1 nm. This is a large distance to be considered for regulatory enhancer-promoter interactions. In contrast, our model predicts an average distance of 68 nm for the interaction between HS40 and α-globin gene in K562 cell line (Figure 5D). Overall, by removing the over-constraints of unrealistic bead radii and by improving sampling, our model provides a more accurate description of the ensemble structures of the α-globin locus.

Our results presented here show that non-specific spatial interactions arising from nuclear confinement and excluded volume effect have significant occurrence in 5C measurements, as up to ∼70–90% of 5C interactions can be accounted for when self-avoiding chromatin chains are confined in the available space of the crowded nucleus. To guard against false positives, we focus on long-range interactions detected with the strongest statistical confidence. As this strategy is rather conservative, a portion of the 5C interactions that are temporarily excluded appear subsequently in the constructed 3D ensemble structures, likely resulting from constraints from the stronger interactions and the confinement of the self-avoiding chromatin chains. While the reappearance of such interactions in our model does not guarantee that we can detect all functionally relevant moderate or weak interactions, we recognize that their detection with high precision is an overall challenging task in the field. Furthermore, our approach is not overly sensitive to the choice of parameters such as the diameter of the spherical confinement. While recent studies shows that the spatial confinement is a major determinant of the genome organization (22,40), as much of the scaling rules including the exponent of looping probability can be explained by the confinement of the self-avoiding chromatin chains, the scaling exponent α changes only slowly as the nuclear diameter in the relevant size regime changes (40). Therefore with stringent False Discovery Rate, moderate changes in nuclear diameter and deviations from the spherical shape are unlikely to affect the identification of the most significant 5C interactions.

Our model reveals that even though the chromatin chains adopt varying degrees of compactness, important interactions are consistently observed. These interactions appear broadly in the configurations of chromatin chains in the active cell line. This finding suggests a common detailed structural scaffold in the active cell line that is required for α-globin expression. The nC-SAC model further allows structural examination of subpopulations of chromatin chains adopting different configurations. As demonstrated by recent single cell studies, cells with identical hormonal stimulation may exhibit diverse levels of gene expression, highly expressed genes at the population level may exhibit bimodal distributions, and epigenetic modifications may be highly heterogeneous (56–58). Access to 3D chromatin structures of subpopulations of cells will help to gain understanding of the structural diversity of chromatin chains associated with the heterogeneity of gene expression and epigenetic modifications.

Our method can make many detailed predictions of spatial interactions between distant genomic elements, some of which are validated by available ChIA-PET studies. Excluding locations that lack 5C coverage or locations where ChIA-PET and 5C measurements disagree, our model recovered all remaining RNAPII, CTCF, and RAD21 mediated long-range chromatin interactions, as well as interactions associated with concurrent histone modifications. While we cannot extrapolate to declare all novel interactions predicted by our model are biologically important, the overall validation by ChIA-PET suggests that our method can make detailed predictions that are biologically relevant. Additional experimental investigations with higher resolution (<100 kb) than 3D-FISH analysis are required for further validation of our predictions.

Our method can also suggest highly specific and testable mechanistic models of gene regulation. While 5C measurement has identified many important chromatin interactions, details of our predicted chromatin chains suggest a complex many-body mechanism of gene regulation that is beyond a simple gene-enhancer model. Although the α-globin gene and the enhancers HS40/46/48 interact in both cell lines, the enhancers interact strongly with POL3RK in the silent but not in the active cells. As POL3RK is observed to have bound transcription factors, we speculate it may occlude access of enhancers to factors necessary for α-globin activation in the silent cells. This mechanism of gene inactivation through denial-of-access is also consistent with the epigenetic profiles of the enhancers and the POL3RK gene in both cell lines. Analogous to the mechanism of a multi-gene complex for co-transcription, in which the promoter of the first gene acts as an enhancer of the second gene (38), a multi-gene complex for inactivation may be at play. Since the accessibility of transcription factor binding is a key determinant of gene regulation (59), the POL3RK gene in this case may act on the enhancers of α-globin gene as a silencer through denial of access of transcription factors. Although these predictions are rather speculative, they can be tested by genetic perturbation of the identified multi-body structural unit. While recent Hi-C studies (60) can identify chromatin interactions at high resolution, the discovery of this many-body mechanism would not be possible without constructing 3D ensemble of structures due to the pairwise nature of Hi-C technique. Further comparison of our results with the high-resolution (5 kb) Hi-C data showed that 66% and 65% of predicted interactions are found to be statistically significant in Hi-C study of GM12878 and K562 cell lines, respectively. This level of agreement may indicate that our nC-SAC method can leverage low resolution experimental observation and build structures compatible with higher resolution observations. The importance of 3D model of chromatin interactions was also demonstrated in a recent study, where a many-body interactions between Sox9 and Kcnj2 genes were discovered (18).

Our study also suggests that integrating 3D models of chromatin chains with epigenetic data can reveal mechanistic insight into the regulation of cell activities. While genome-wide epigenetic studies such as CTCF enrichment and histone modification point to potential regulatory elements and suggest possible long-range interactions along the one-dimensional genome (53), it is challenging to interpret and integrate such information. Recent studies showed that important organizational properties of genome such as the formation of TADs can be inferred from the integration of epigenome data with 3D structure construction (15–17). Along this line, by projecting epigenetic data onto predicted 3D chromatin chains, here we have shown that one can gain better understanding of the complex many-body machineries of gene regulation that involves multiple genomic elements.

Our method is general and can be applied to determine configurations of other gene loci. However, successful predictions are limited by the availability, consistency, and resolution of experimental measurements. In addition, while our method can predict novel interactions, such predictions can only be made in neighborhoods with rich contact information. As the density of experimentally captured interactions decreases, successful predictions become less likely. In principle, any 3C and related data (4C/5C/Hi-C) can be used as spatial constraints to infer 3D chromatin ensembles. With additional algorithm development, the nC-

SAC method can be further improved so it can generate 3D ensembles of chromatins from high resolution Hi-C data. To incorporate Hi-C data, further improvement in sampling with more elaborated techniques of resampling and rejection control will be necessary to allow studies of longer chromatin chains while incorporating more spatial constraints. In summary, the nC-SAC method can model chromatin structures of gene loci in cell populations and subpopulations with different expression levels. It also provides a new approach for identifying spatial structures and interactions and for assessing their roles in regulating gene activities. These results point to exciting opportunities of leveraging limited and pairwise chromosome conformation capture data through modeling of 3D chromatin structures to gain additional knowledge on long-range interactions. Combined with further genetic manipulation, we expect future studies will lead to novel insight into the spatial organization of the genome.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank H. Naveed and R. Wuerffell for valuable discussions and M. Macellaio for reading the manuscript.

## FUNDING

National Institutes of Health [R01GM126558 to J.L., R2 1AI126308 to J.L. and A.K., R01AI121286, R21AI117687 to A.K.]; Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust (to A.K. and J.L.). Funding for open access charge: Research Open Access Publishing (ROAAP) Fund of the University of Illinois at Chicago.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Fraser,P. and Bickmore,W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature*, **447**, 413–417.
2. Rippe,K., Dekker,M., Dekker,J. and Kleckner,N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.
3. Hagège,H., Klous,P., Braem,C., Splinter,E., Dekker,J., Cathala,G., de Laat,W. and Forné,T. (2007) Quantitative analysis of chromosome conformation capture assays (3c-qpcr). *Nat. Protoc.*, **2**, 1722–1733.
4. Miele,A., Bystricky,K. and Dekker,J. (2009) Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. *PLoS Genet.*, **5**, e1000478.
5. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
6. Duan,Z., Andronescu,M., Schutz,K., McIlwain,S., Kim,Y.J., Lee,C., Shendure,J., Fields,S., Blau,C.A. and Noble,W.S. (2010) A three-dimensional model of the yeast genome. *Nature*, **465**, 363–367.
7. Montefiori,L., Wuerffel,R., Roqueiro,D., Lajoie,B., Guo,C., Gerasimova,T., De,S., Wood,W., Becker,K.G., Dekker,J. *et al.* (2015) Extremely long-range chromatin loops link topological domains to facilitate a diverse antibody repertoire. *Cell Rep.*, **14**, 896–906.
8. Nora,E.P., Lajoie,B.R., Schulz,E.G., Giorgetti,L., Okamoto,I., Servant,N., Piolot,T., van Berkum,N.L., Meisig,J., Sedat,J. *et al.* (2012) Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, **485**, 381–385.
9. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
10. Phillips-Cremins,J.E., Sauria,M.E., Sanyal,A., Gerasimova,T.I., Lajoie,B.R., Bell,J.S., Ong,C.T., Hookway,T.A., Guo,C., Sun,Y. *et al.* (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, **153**, 1281–1295.
11. Zhang,Y., Dudko,O.K., Lucas,J.S. and Murre,C. (2014) 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell*, **158**, 339–352.
12. Ay,F. and Noble,W.S. (2015) Analysis methods for studying the 3D architecture of the genome. *Genome Biol.*, **16**, 183.
13. Junier,I., Spill,Y.G., Marti-Renom,M.A., Beato,M. and le Dily,F. (2015) On the demultiplexing of chromosome capture conformation data. *FEBS Lett.*, **589**, 3005–3013.
14. Di Pierro,M., Zhang,B., Aiden,E.L., Wolynes,P.G. and Onuchic,J.N. (2016) Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 12168–12173.
15. Junier,I., Dale,R.K., Hou,C., Kepes,F. and Dean,A. (2012) CTCF-mediated transcriptional regulation through cell type-specific chromosome organization in the β-globin locus. *Nucleic Acids Res.*, **40**, 7718–7727.
16. Brackley,C.A., Johnson,J., Kelly,S., Cook,P.R. and Marenduzzo,D. (2016) Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.*, **44**, 3503–35012.
17. Jost,D., Carrivain,P., Cavalli,G. and Vaillant,C. (2014) Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.*, **42**, 9553–9561.
18. Chiariello,A.M., Annunziatella,C., Bianco,S., Esposito,A. and Nicodemi,M. (2016) Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.*, **6**, 29775.
19. Giorgetti,L., Galupa,R., Nora,E.P., Piolot,T., Lam,F., Dekker,J., Tiana,G. and Heard,E. (2014) Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, **157**, 950–963.
20. Tokuda,N., Terada,T.P. and Sasai,M. (2012) Dynamical modeling of three-dimensional genome organization in interphase budding yeast. *Biophys J.*, **102**, 296–304.
21. Barbieri,M., Chotalia,M., Fraser,J., Lavitas,L.M., Dostie,J., Pombo,A. and Nicodemi,M. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *PNAS*, **109**, 16173–16178.
22. Kang,H., Yoon,Y., Thirumalai,D. and Hyeon,C. (2015) Confinement-induced glassy dynamics in a model for chromosome organization. *Phys. Rev. Lett.*, **115**, 198102.
23. Goloborodko,A., Marko,J.F. and Mirny,L.A. (2016) Chromosome compaction by active loop extrusion. *Biophys. J.*, **110**, 2162–2168.
24. Fudenberg,G., Imakaev,M., Lu,C., Goloborodko,A., Abdennur,N. and Mirny,L.A. (2016) Formation of chromosomal domains by loop extrusion. *Cell Rep.*, **15**, 2038–2049.
25. 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. *et al.* (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **12**, E6456–E6465.
26. Zhang,B. and Wolynes,P.G. (2015) Topology, structures, and energy landscapes of human chromosomes. *PNAS*, **112**, 6062–6067.
27. Tjong,H., Gong,K., Chen,L. and Alber,F. (2012) Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.*, **22**, 1295–1305.
28. Wong,H., Marie-Nelly,H., Herbert,S., Carrivain,P., Blanc,H., Koszul,R., Fabre,E. and Zimmer,C. (2012) A predictive computational model of the dynamic 3D interphase yeast nucleus. *Curr Biol.*, **22**, 1881–1890.
29. Baù,D., Sanyal,A., Lajoie,B.R., Capriotti,E., Byron,M., Lawrence,J.B., Dekker,J. and Marti-Renom,M.A. (2010) The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.*, **18**, 107–114.

30. Fraser,J., Ferraiuolo,M.A., Dostie,J., Blanchette,M. and Rousseau,M. (2011) Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinformatics*, **12**, 414.
31. Kalhor,R., Tjong,H., Jayathilaka,N., Alber,F. and Chen,L. (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.*, **30**, 90–98.
32. Meluzzi,D. and Arya,G. (2013) Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Res.*, **41**, 63–75.
33. Ay,F., Bunnik,E.M., Varoquaux,N., Bol,S.M., Prudhomme,J., Vert,J.P., Noble,W.S. and Le Roch,K.G. (2014) Three-dimensional modeling of the P. Falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.*, **24**, 974–988.
34. Trieu,T. and Cheng,J. (2014) Large-scale reconstruction of 3D structures of human chromosomes from chromosomal contact data. *Nucleic Acids Res.*, **42**, e52.
35. Wang,S., Xu,J. and Zeng,J. (2015) Inferential modeling of 3D chromatin structure. *Nucleic Acids Res.*, **43**, e54.
36. Tjong,H., Li,W., Kalhor,R., Dai,C., Hao,S., Gong,K., Zhou,Y., Li,H., Zhou,X.J., Le Gros,M.A. *et al.* (2016) Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *PNAS*, **113**, E1663–E1672.
37. Belmont,A.S. (2014) Large-scale chromatin organization: the good, the surprising, and the still perplexing. *Curr. Opin. Cell Biol.*, **26**, 69–78.
38. Li,G., Ruan,X., Auerbach,R.K., Sandhu,K.S., Zheng,M., Wang,P., Poh,H.M., Goh,Y., Lim,J., Zhang,J. *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, **148**, 84–98.
39. Heidari,N., Phanstiel,D.H., He,C., Grubert,F., Jahanbani,F., Kasowski,M., Zhang,M.Q. and Snyder,M.P. (2014) Genome-wide map of regulatory interactions in the human genome. *Genome Res.*, **24**, 1095–1917.
40. Gürsoy,G., Xu,Y., Kenter,A.L. and Liang,J. (2014) Spatial confinement is a major determinant of folding landscape of human chromosomes and genetic programming of cell. *Nucleic Acids Res.*, **42**, 8223–8230.
41. Boettiger,A.N., Bintu,B., Moffitt,J.R., Wang,S., Beliveau,B.J., Fudenberg,G., Imakaev,M., Mirny,L.A., Wu,C.T. and Zhuang,X. (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, **529**, 418–422.
42. Naumova,N., Smith,E.M., Zhan,Y. and Dekker,J. (2012) Analysis of long-range chromatin interactions using chromosome conformation capture. *Methods*, **58**, 192–203.
43. Yoav,B. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, **57**, 289–300.
44. Liang,J., Zhang,J. and Chen,R. (2002) Statistical geometry of packing defects of lattice chain polymer from enumeration and sequential Monte Carlo method. *J. Chem. Phys.*, **117**, 3511–3521.
45. Zhang,J., Chen,R., Tang,C. and Liang,J. (2003) Origin of scaling behavior of protein packing density: A sequential Monte Carlo study of compact long chain polymers. *J. Chem. Phys.*, **118**, 6102.
46. Lin,M., Lu,H.M., Chen,R. and Liang,J. (2008) Generating properly weighted ensemble of conformations of proteins from sparse or indirect distance constraints. *J. Chem. Phys.*, **129**, 094101.
47. Lin,M., Chen,R. and Liang,J. (2008) Statistical geometry of lattice chain polymers with voids of defined shapes: sampling with strong constraints. *J. Chem. Phys.*, **128**, 084903.
48. Zhang,J., Dundas,J., Lin,M., Chen,R., Wang,W. and Liang,J. (2009) Prediction of geometrically feasible three dimensional structures of pseudoknotted RNA through free energy estimation. *RNA*, **15**, 2248–2263.
49. Cao,Y. and Liang,J. (2013) Adaptively biased sequential importance sampling for rare events in reaction networks with comparison with exact solutions from finite buffer dCME method. *J. Chem. Phys.*, **139**, 025101.
50. Lin,M., Chen,R. and Liu,J.S. (2013) A lookahead strategies for sequential Monte Carlo. *Stat. Sci.*, **28**, 69–94.
51. Zhou,G.L., Xin,L., Song,W., Di,L.J., Liu,G., Wu,X.S., Liu,D.P. and Liang,C.C. (2006) Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes. *Mol. Cell. Biol.*, **26**, 5096–5105.
52. Vernimmen,D., Marques-Kranc,F., Sharpe,J.A., Sloane-Stanley,J.A., Wood,W.G., Wallace,H.A., Smith,A.J. and Higgs,D.R. (2009) Chromosome looping at the human α-globin locus is mediated via the major upstream regulatory element (hs-40). *Blood*, **114**, 4253–4260.
53. Zeng,C., Guo,X., Long,J., Kuchenbaecker,K.B., Droit,A., Michailidou,K., Ghoussaini,M., Kar,S., Freeman,A., Hopper,J.L. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
54. Ester,M., Kriegel,H.P., Sander,J. and Xu,X. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, 226–231.
55. Doyle,B., Fudenberg,G., Imakaev,M. and Mirny,L.A. (2014) Chromatin loops as allosteric modulators of enhancer-promoter interactions. *PLoS Comput. Biol.*, **10**, 1003867.
56. Shalek,A.K., Satija,R., Shuga,J., Trombetta,J.J., Gennert,D., Lu,D., Chen,P., Gertner,R.S., Gaublomme,J.T., Yosef,N. *et al.* (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*, **510**, 363–369.
57. Shalek,A.K., Satija,R., Adiconis,X., Gertner,R.S., Gaublomme,J.T., Raychowdhury,R., Schwartz,S., Yosef,N., Malboeuf,C., Lu,D. *et al.* (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, **498**, 236–240.
58. Rotem,A., Ram,O., Shoresh,N., Sperling,R.A., Goren,A., Weitz,D.A. and Bernstein,B.E. (2015) Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.*, **33**, 1165–1172.
59. Fraser,P. (2006) Transcriptional control thrown for a loop. *Curr. Opin. Genet. Dev.*, **16**, 490–495.
60. Rao,S.S., Huntley,M.H., Durand,N.C., Stamenova,E.K., Bochkov,I.D., Robinson,J.T., Sanborn,A.L., Machol,I., Omer,A.D., Lander,E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.