Methods 58 (2012) 300–306

Contents lists available at SciVerse ScienceDirect

Methods

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

Review Article

Genome structure determination via 3C-based data integration by the Integrative Modeling Platform

Davide Baù, Marc A. Marti-Renom *

Structural Genomics Team, Genome Biology Group, National Center for Genomic Analysis (CNAG), Barcelona, Spain

ARTICLE INFO

Article history:
Available online 13 April 2012

Keywords:
Chromosome conformation capture
Chromatin structure
Integrative modeling
Computational structural biology

ABSTRACT

The three-dimensional (3D) architecture of a genome determines the spatial localization of regulatory elements and the genes they regulate. Thus, elucidating the 3D structure of a genome may result in significant insights about how genes are regulated. The current state-of-the art in experimental methods, including light microscopy and cell/molecular biology, are now able to provide detailed information on the position of genes and their interacting partners. However, such methods by themselves are not able to determine the high-resolution 3D structure of genomes or genomic domains. Here we describe a computational module of the Integrative Modeling Platform (IMP, http://www.integrativemodeling.org) that uses chromosome conformation capture data to determine the 3D architecture of genomic domains and entire genomes at unprecedented resolutions. This approach, through the visualization of looping interactions between distal regulatory elements, allows characterizing global chromatin features and their relation to gene expression. We illustrate our work by outlining the determination of the 3D architecture of the α-globin domain in the human genome.

1. Introduction

The knowledge of the three-dimensional (3D) architecture of a genome, or a genomic domain, is essential for characterizing how genes and their regulatory elements get spatially close to carry out their function [1]. Previous attempts to model the conformation of chromatin using computational approaches include polymer physics [2,3] and molecular dynamics [4]. These methods have proved useful in providing insights into chromatin flexibility, compaction and unpacking [5–8]. However, until recently, computational modeling did not take full advantage of the experimental data available on chromatin folding. At present, the highest resolution data on genome architectures is mainly available by light microscopy technologies [9] and the so-called chromosome conformation capture (3C)-based approaches [2,10–13]. In particular, the 3C Carbon Copy technology (5C) was developed to allow the simultaneous detection of interactions within a genomic domain or even entire chromosomes [14,15], in contrast to the original 3C technique in which only single pair-wise loci could be investigated at a time [2]. In 5C, the PCR step of 3C is replaced by ligation-mediated amplification (LMA) followed by the detection of ligation products. With LMA it is possible to use simultaneously thousands of primers, allowing the parallel detection of millions of chromatin interactions. 5C experiments result in a matrix of interaction frequencies between loci located within the studied genomic domain. Although it has been shown that chromatin interaction frequencies can be used as a proxy for spatial distance between interacting fragments [10], they do not give direct information on the 3D organization of the corresponding interacting loci. Therefore, the integration of 5C-based experiments with computational methods becomes essential for determining the 3D conformation of a genomic domain.

With the availability of chromatin interaction data, new methods aiming at resolving the 3D folding of a genomic domain or entire genomes at medium to high resolution by integrating experimental data into computational methods have been developed [16–20]. Such works demonstrate that integrating experimental data from different sources can help to overcome the limitations of individual experimental techniques for determining the spatial organization of chromatin. For example, using multiple data sources carrying different information can improve the final resolution of the models by increasing the individual signal to noise ratio. Here we describe a method to generate medium to high-resolution models of chromatin regions by integrating 5C experimental data into the Integrative Modeling Platform (IMP, http://www.integrativemodeling.org) [21]. We show that this approach can be used to determine the 3D architecture of chromatin at unprecedented resolutions, as presented with the modeling of the 3D architecture of the α-globin domain in chromosome 16 of

* Corresponding author. Address: Structural Genomics Team, Genome Biology Group, National Center for Genomic Analysis (CNAG), c/Baldiri Reixac, 4 PCB – Tower I, 2nd floor, 08028 Barcelona, Spain. Fax: +34 934 037 279.
E-mail address: mmarti@pcb.ub.cat (M.A. Marti-Renom).

1046-2023/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2012.04.004

the human genome, which has been recently described in detail [18]. The method reviewed here improved the resolution of the models generated in previous works [16,19], which were limited by the resolution and completeness of the input experimental data, by insufficient model representation, scoring and optimization, or by limited analysis of the 3D models.

2. Structure determination by IMP

The IMP conceptual framework consists of four steps: representation, scoring, optimization, and analysis.

2.1. Representation

The system representation in IMP (*i.e.,* how the biomolecule under investigation is represented in the IMP framework) must be detailed enough to comprehensively represent the experimental data without making the search of the 3D conformational space unfeasible due to an excess of complexity. The system representation details (or resolution) determine the accessible conformational space. While coarse-grained representations are more suitable for the analysis of large conformational searches (*i.e.,* complex systems like biological macromolecules), fine-grained representations require more computational power to explore the same search space. IMP represents an object by a set of hierarchical particles and their properties (including their Cartesian coordinates that define their spatial position). The particles hierarchy allows for a flexible representation of the system at different resolutions matching the appropriate representation of the diverse input data.

2.2. Scoring

The key step in structure determination methods is the proper evaluation of the generated models by assigning a score to each model. Such score will allow identifying the different solutions that are compatible with the input experimental data. Therefore, a measurable relationship between the particles representing the physical system and the system itself must be formulated. In IMP this is accomplished by a joint probability density function (pdf) that affect the attributes of a particle, including their spatial coordinates. A pdf is a function that describes the probability of a variable (e.g., the coordinates of each particle of the model) to take on a given value (e.g., the distance between two particles derived from the experimental input data). Each independent observation of the system results in a number of pdf related to a particle or to a group of particles. That is, each pdf may affect a single particle or a set of them. The final scoring function, also called IMP objective function, will result from the sum of each individual pdf affecting all particles in the system. It is important to note that not all the individual pdfs will have the same weight and thus their violation may account differently to the final IMP objective function. The functional forms of the restraints in IMP are diverse [22] and were initially implemented to determine the structure of protein assemblies [23].

2.3. Optimization

Once the system has been appropriately represented and the relationship between the constituent particles has been formulated in the form of a scoring function, IMP searches for a conformational solution of the modeled object by minimizing its IMP objective function (*i.e.,* simultaneously reducing the violations of all the imposed restraints). The conformational space of chromatin is large and may have numerous local minima depending on the amount and type of imposed restraints from experiments. Therefore, to comprehensively explore the conformation space is not a straightforward task, which can be addressed by the battery of IMP optimization protocols (such as Monte Carlo method). Even though it is computationally impossible to enumerate all possible solution of the conformational space, our protocol accomplished the search by generating a large number of the most common solutions (*i.e.,* 3D solution that best satisfies the imposed restraints). The selection of the optimization protocol in IMP will depend on the representation and scoring schema of the system.

2.4. Analysis

Finally, the structural analysis of the resulting ensemble of solutions consistent with the input restraints (*i.e.,* input experimental data) aims at validating the models and revealing important aspects of the modeling protocol. Moreover, such analysis may results in observations that could be useful for the re-designing of new experiments, which, in turn, could lead to higher resolution models. The final step is thus essential for leveraging the iterative nature of the integrative modeling.

A general description of the IMP framework was recently published [21]. Here, we focus on delineating all the necessary steps for generating 3D models of a genomic domain based on 5C experimental data integration into IMP.

3. Modeling genomic domains with IMP

3.1. 5C data normalization

5C experimental data consist of frequencies of interaction between restriction fragments, which can be considered as a proxy for the spatial vicinity of two fragments. In order to normalize possible biases in the experimental data, an internal normalization by means of Z-scoring the frequency data is applied. Generally, a Z-score indicates how many standard deviations an experimental observation is above or below the mean of the data. This normalization allows quantifying the variability within the 5C matrix and identifying pairs of fragments that interact above or below the average 5C interaction frequency. However, the Z-score calculation requires the input data to follow a normal distribution centered on its average. Since the raw 5C counts do not follow a normal distribution, values are transformed by applying a $\log_{10}$ to the raw experimental counts and then checked for their normality. The Z-score of the $\log_{10}$ value of the raw frequency ($c_{i,j}$) for the interacting fragments i and j is then computed as:

$$Zscore_{i,j} = \frac{(c_{i,j} - \mu)}{\sigma} \quad (1)$$

where i and j, are restriction fragments and μ and σ are the average and standard deviation of $c_{i,j}$ of the entire 5C matrix.

3.2. Model representation and scoring function

IMP assigns a particle to each restriction fragment resulting from the 5C experiment, which is represented in the 3D space as a sphere of radius proportional to the number of nucleotides in the fragment (or nucleotide length l). An excluded volume proportional to l is set to avoid clashes between particles. Assuming a canonical 30 nm fiber, for a particle i of nucleotide length l_i the excluded volume is defined by a radius r_i as:

$$r_i = 0.005 \cdot l_i \quad (2)$$

The spatial position of the particles representing the studied chromatin region is determined from the input 5C experimental data by satisfying series of restraining oscillators (or springs) implemented between each particle pair. The objective of the restraining oscillators is to maintain a particle pair at a given

equilibrium distance derived from the corresponding 5C Z-score. Equilibrium distances are inversely proportional to the 5C Z-scores. However, since the exact form of this proportion is unknown, we have implemented two linear relationships to map 5C Z-scores onto their corresponding equilibrium distances. The implementation of two different relationships was needed because of the different response to 5C experiments between sequentially close fragments and other pairs of fragments [24]. This difference is due to a vicinity effect of the experimental technique for which sites sequentially separated by few kilobases (kb) interact much more frequently than sites separated by tens or hundreds of kb. The two linear relationships between 5C Z-scores and spatial distances are defined by: (i) two neighbor fragments (*i.e.,* fragments i to $i+1\cdots2$) are restrained based on the linear relationship between the 5C Z-scores and the sum of their excluded volume, calculated as in Eq. (2), and (ii) two non-neighbor fragments (*i.e.,* fragments i to $i+3\cdots n$) are restrained based on the relationship between the Z-scores and optimized distance values (see Section 3.3). These two linear relationships rely on the assumptions that: (i) consecutive fragments are spatially restrained proportionally to their excluded volume upon the relationship of 0.01 nm per base pair, assuming a canonical 30 nm fiber [25]; and (ii) two non-neighbor fragments cannot get closer than 30 nm, which corresponds to the diameter of the chromatin fiber [26]. Finally, the values of two Z-scores cut-offs have to be set to define the type of restraint imposed between two non-neighbor fragments (see also Section 3.3).

Once the distances between the particles have been defined, a solution is obtained by searching for a 3D representation of the system that best satisfies all the imposed restraints. The spatial position of each particle is determined by satisfying series of restraining oscillators (or springs) implemented between pairs of particles, which aim at maintaining them at a given equilibrium distance. Three types of restraining oscillators are used for modeling: (i) harmonic oscillators ($H_{i,j}$), which aim at positioning a pair of particles at a given equilibrium distance; (ii) lower-bound harmonic oscillators ($lbH_{i,j}$), which avoid two particles to get closer than a given equilibrium distance and; (iii) upper-bound harmonic oscillators ($ubH_{i,j}$), which force two particles to be closer than a given equilibrium distance. The exact functions of the restraints are:

$$H_{i,j} = k(d_{i,j} - d^0_{i,j})^2 \tag{3}$$

$$\begin{cases} if\ d_{i,j} \leqslant d^0_{i,j}; & lbH_{i,j} = k(d_{i,j} - d^0_{i,j})^2 \\ if\ d_{i,j} > d^0_{i,j}; & lbH_{i,j} = 0 \end{cases} \tag{4}$$

$$\begin{cases} if\ d_{i,j} \geqslant d^0_{i,j}; & ubH_{i,j} = k(d_{i,j} - d^0_{i,j})^2 \\ if\ d_{i,j} < d^0_{i,j}; & ubH_{i,j} = 0 \end{cases} \tag{5}$$

where $d_{i,j}$ is the current distance between particles i and j during simulation, $d^0_{i,j}$ is the equilibrium distance obtained by mapping 5C Z-scores onto distances, and k is the force constant applied to the restraint, which scales the penalty added to the IMP objective function for not satisfying a restraint as the square root of the absolute value of the 5C Z-score between the two interacting particles. The force constant was implemented in this way so that to give more relevance to extreme positive and negative 5C Z-scores values (*i.e.,* the more an experimental data differs from the mean value, the higher its weight is). An exception to this rule is applied to neighbor fragments, for which the force is set to an arbitrary value large enough to maintain connectivity between neighbor fragments.

The type of restraint (*i.e.,* $H_{i,j}$, $lbH_{i,j}$, or $ubH_{i,j}$) and the equilibrium distance applied to each particle are defined based on the 5C experimental data and three IMP parameters: (i) a lower-bound Z-score cut-off (*lZ*), (ii) a upper-bound Z-score cut-off (*uZ*), and (iii) a maximal proximity for two non-interacting fragments (*mP*). The optimal values for these three parameters are empirically determined for each set of experimental data via a calibration protocol (see Section 3.4). Interaction Z-scores between the *lZ* and *uZ* cut-offs, which correspond to Z-scores close to the average interaction frequencies, are not used during modeling by IMP. Using exclusively Z-scores below *lZ* and above *uZ* restricts the modeling to the use of restraints from pairs of interacting fragments with very low and very high interaction frequencies, respectively. Finally, the *mP* parameter sets the minimum distance between two pairs of non-interacting fragments (*i.e.,* 5C interaction frequency of zero). These three parameters need to be determined empirically for each set of experimental data, as described below.

3.3. IMP calibration

Equilibrium distances are inversely proportional to the 5C Z-scores. Two different linear relationships have to be defined for neighbor (*i.e.,* fragments i to $i+1\cdots2$) and non-neighbor (*i.e.*, fragments i to $i+3\cdots n$) fragments. Neighbor fragments are set at an equilibrium distance proportional to the sum of their occupied excluded volume. The sum of radii of each pair of neighbor particles (calculated using Eq. (2)) is plot against their corresponding Z-score value. The equation of the linear interpolation of this plot is then used to compute the equilibrium distances for neighbor fragments (Fig. 1c, red linear fitting).

For non-neighbor fragments, the equation of the line joining the two points defined by the maximum Z-score value with the closest distance between two condensed chromatin fragments (30 nm) (first point), and the minimum Z-score value with the *mP* parameter (second point), is used as the distance mapping function (Fig. 1c, yellow line).

The type of harmonic restraint applied to a pair of particles depends on whether the pairs of particles are neighbors or non-neighbors and on the *lZ* and *uZ* values.

3.4. Empirical determination of IMP parameters

The values of the parameters *mP*, *lZ* and *uZ* to be used for modeling the genomic domain under investigation are empirically determined over a grid search, exploring various combinations of different ranges of *mP* values around the size of the largest pair of restriction fragments left after digestion, with *lZ* and *uZ* cut-off values in bins of 0.1 ranging from −1.0 to 1.0. The procedure is as follow: (i) for each set of parameters, 500 models are generated using the protocol described in the section *Model building with IMP*; (ii) a contact map, which can be regarded as a computational 5C interaction matrix, is calculated from the 500 models where a contact corresponds to two particles within 200 nm; and (iii) the correlation coefficient between the calculated contact map and the experimental 5C matrix is computed. The optimal *mP*, *lZ* and *uZ* values correspond to the grid cell with the maximum correlation coefficient between the contact map and the experimental 5C matrix. Such selection of parameters ensures that the resulting 3D models best fit the input 5C raw data. Ideally, the correlation coefficient between the two matrices (*i.e.,* the experimental 5C counts and the 3D models contact maps) would be near 1.0, indicating that the resulting ensemble of models explains most, if not all, the input 5C data. However, 5C experiments capture the macroscopic state of chromatin in a population of cells (*i.e.,* it is population-averaged) and thus, the resulting correlation coefficient is always lower than 1.0 since none of the individual models is likely to reflect all the contacts in the input matrix.

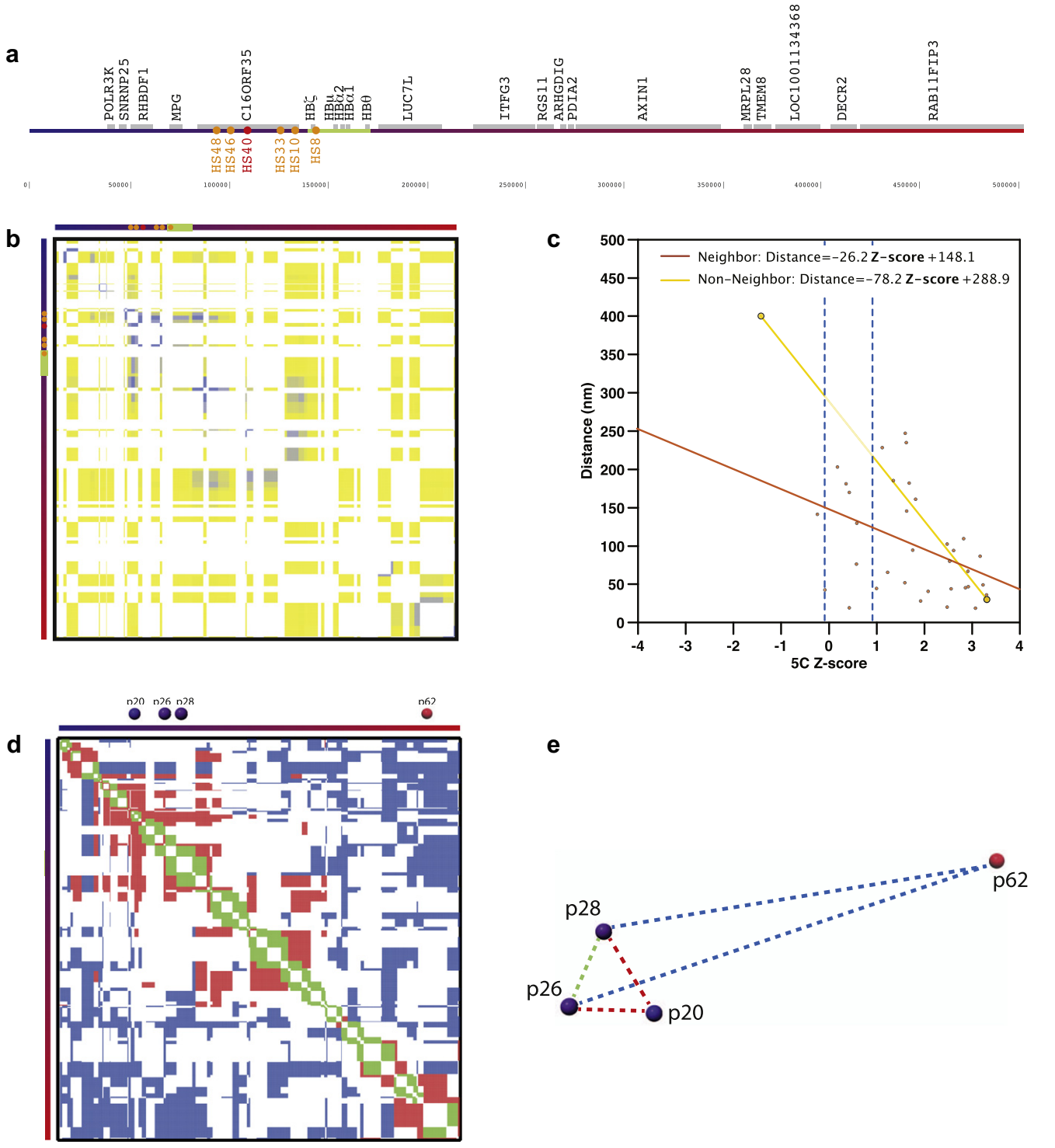

Fig. 1. Structure determination of the α-globin domain. (a) Linear (1D) map of the 0.5 Mb ENm008 region including the ζ, μ, α2, α1, and θ globin genes. Gray lines above the linear representation colored from blue (telomere) to red (centromere) show annotated genes in the region. Red and orange circles localize the HS40 and other α-globin related HS sites. (b) 5C Z-scores matrix. Blue to yellow color indicate positive to negative Z-scores. White color indicates non-interrogated pairs. For easy inspection, the 1D representation of the ENm008 is used in the *x* and *y* axis of the plot. (c) IMP calibration curves to assess the linear relationship between 5C Z-scores and equilibrium distances between neighbor (red linear fitting) and non-neighbor (yellow line) fragments. Two vertical dashed blue lines indicate the optimal Z-scores upper- and lower-cut-offs. (d) Harmonic (red), lower-bound harmonic (blue) and upper-bound harmonic (green) restraints applied to the pairs of restrained fragments during simulation. For easy inspection, the 1D representation of the ENm008 is used in the *x* and *y* axis of the plot. (e) Schematic representation of four particles (numbers 20, 26, 27 and 62) and the applied restraints between them, which are represented as a dashed line colored as the restraint type. For easy inspection, the particles 1D position is also shown in panel d.

3.5. Restraints definition

Once the three parameters *mP*, *lZ* and *uZ* have been defined, the type of restraints between each pair of particles can be set. First, two neighbor particles with calculated 5C Z-scores are restrained by a harmonic oscillator with an equilibrium distance proportional to their 5C Z-score following the neighbor linear relationship (Fig. 1c, red linear fitting). Due to the presence of repetitive elements in

the genome, it might be possible that some restriction fragments were not interrogated in the 5C analysis for the unavailability of unique 5C primers. Therefore, two neighbor particles with no calculated 5C Z-scores are restrained by an upper-bound harmonic oscillator with an equilibrium distance corresponding to their excluded volume (*i.e.*, as calculated in Eq. (2)). A *k* force higher than the ones for non-neighbor restraints is applied to ensure connectivity between all neighbor fragments. Second, two non-neighbor particles with calculated 5C Z-scores are modeled at a distance and force proportional to their corresponding 5C Z-scores following the non-neighbor linear relationship as described in the section 3.3 (Fig. 1c, yellow linear fitting). Pairs of non-neighbor particles with Z-score higher than the upper-bound cut-off (*uZ*) are restrained by a harmonic oscillator, while pairs of non-neighbor particles with Z-score lower than the lower-bound cut-off (*lZ*) are restrained by a lower-bound harmonic oscillator. Hence, pairs of non-neighbor particles with a Z-score value above the *uZ* parameter are kept close in space, while pairs of non-neighbor particles with a Z-score value below the *lZ* parameter are kept apart in space. The *k* force applied to these restraints corresponds to the square root of the absolute value of their corresponding Z-scores. If for a pair of non-neighbor particles no 5C Z-score is available, they are restrained based on the average 5C-Z-score calculated from the adjacent particles.

3.6. Model building with IMP

IMP generates structures by simultaneously minimizing the violations of all the imposed restraints. Given the dynamic nature of chromatin [27] and the fact that 5C analyses are population-averaged, the optimization of the imposed restraints may result in different configurations with a similar IMP objective function (*i.e.*, different 3D models compatible with the experimental data). Therefore, to comprehensively explore the conformational space of chromatin, the generation of a large number of independent models is needed. For each individual simulation, the IMP building protocol (Fig. 2a) starts by assigning to all particles a set of random Cartesian coordinates within a cube of 1 μm side, which can be exceeded by any particle in the system during the optimization protocol. The optimization is carried out by a combination of 500 Monte Carlo rounds with 5 local steps in a molecular dynamics simulation with a standard simulated annealing method [28]. At each step of the optimization, the current conformation is randomly changed and the change is accepted or rejected according to the Metropolis criteria [29]. The driving scoring function that is minimized during the optimization protocol consists of the sum of all the individual restraint scores between all the particles representing the system.

3.7. Analysis of the generated models

The analysis of the generated models is performed via structural comparison using pair-wise rigid-body superposition that minimizes the root mean square deviation (RMSD) between the superposed conformations [30]. This results in a matrix of all-against-all structural comparisons storing the number of particles for every pair of aligned models that align within 75 nm distance cut-off. The distance cut-off depends on the size of the chromatin region and on the model resolution. This comparison matrix is then used as input file to the Markov Cluster Algorithm (MCL) program [31], which generates unsupervised sets of clusters of related structures. Two main parameters, the pre-inflation (*pi*) and the inflation (*I*), affect the cluster granularity in the MCL program. Similarly to the empirical optimization of IMP parameters, the values

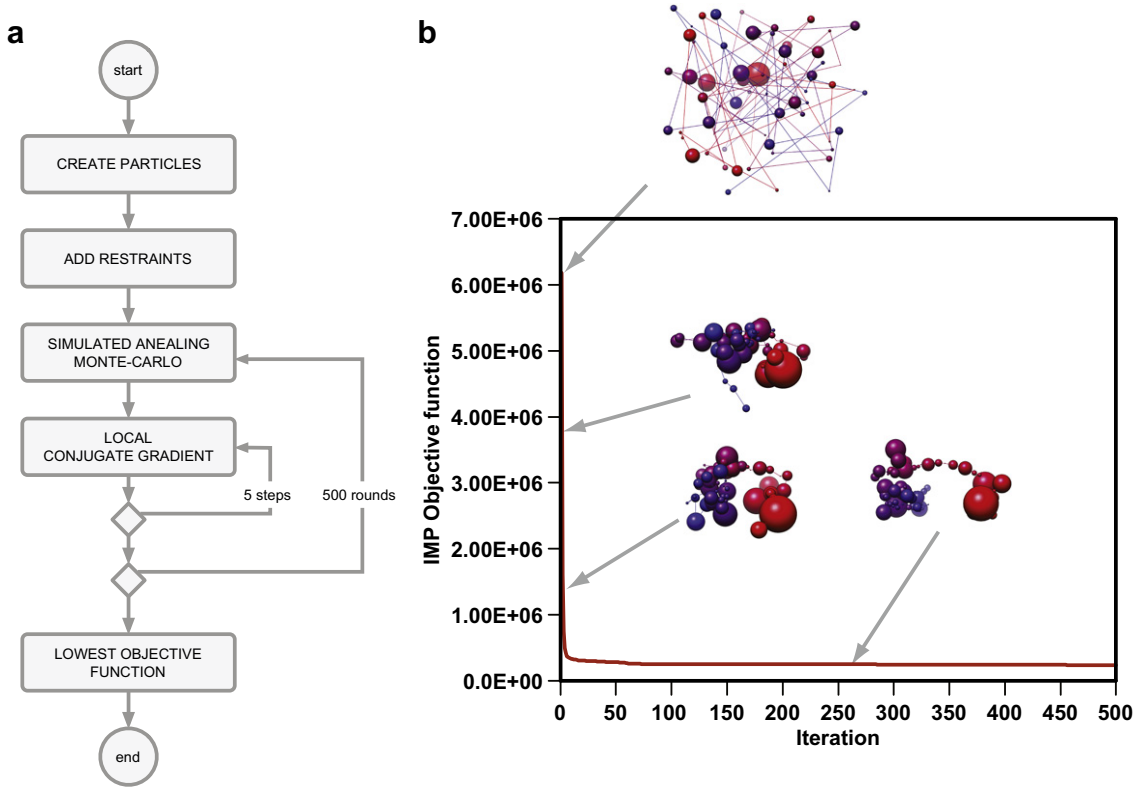

Fig. 2. IMP optimization protocol. (a) Flowchart of the IMP optimization protocol. (b) Schematic representation of a typical optimization process for a single simulation (in the figure a K562 model is displayed). The modeling starts with a random configuration and ends with an optimal configuration after the minimization of the IMP objective function accounting for all violated restraints. Models are show for four different snapshots during the optimization. Each restriction fragment is represented as a particle of radius proportional to its excluded volume.

of these two parameters are set so that the top cluster of models results in the highest correlation between the contact map calculated from models within the cluster and the input 5C experimental matrix. The resulting models in the selected cluster of solutions are ensured to reflect as much as possible the input experimental data, in this case 5C interaction matrices. However, such models need to be independently validated with additional experimental data never used during modeling. Unfortunately, the unavailability of high-resolution experimental techniques for determining the structure of chromatin domains makes it impossible to fully validate the models generated with our method. Therefore, model validation is normally carried out by: (i) mapping onto the 3D models additional one-dimensional data such as epigenetic markers and assessing the biological relevance of the resulting maps; and (ii) using light microscopy techniques such as Fluorescence *In Situ* Hybridization (FISH) [32] to partially validate observed structural features in the models, which can be useful for confirming the global structure of the models.

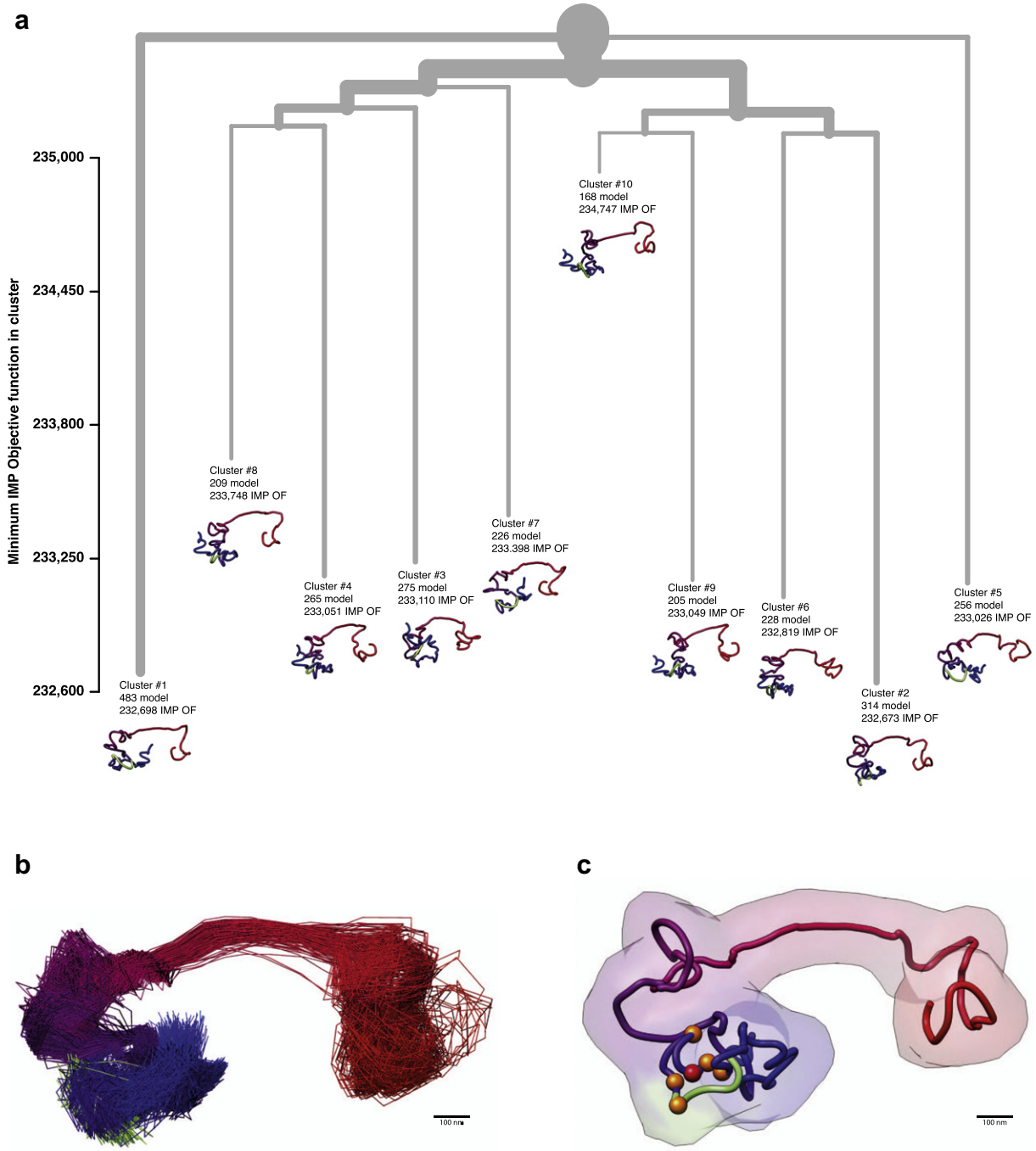

Fig. 3. Ensemble of solutions for the α-globin 3D models. (a) Cluster analysis for the selected 10,000 models showing the structural relationship between the top 10 cluster centroids. The tree was generated based on the structural similarity between each of the centroids. The branch thickness is proportional to the number of solutions at each branch point. Each centroid, colored as in its linear representation (Fig. 1a), is vertically placed proportional to the lowest IMP objective function within the cluster it represents. (b) Ensemble of solutions in cluster number 2. (c) Model representation for the centroid of cluster 2, which is colored as in its linear representation (Fig. 1a) surrounded by a transparent surface calculated with a Gaussian of 175 nm resolution. Images of the shown structures were generated using the Chimera program [38].

4. A proof of principle: the 3D architecture of the human α-globin domain

The described method was recently applied for determining the 3D architecture of the human α-globin domain, a ~500 Kb region located in the human chromosome 16 (Fig. 1a). The α-globin domain has been classically used as a model to study the mechanisms of long-range gene regulation [33–35]. A comprehensive interaction map of the α-globin locus was obtained by 5C experiments [18] (Fig. 1b). 5C experiments were performed on K562 cells, where long-range interactions are expected to occur between the α-globin genes and their regulatory elements. As described above, to each restriction fragment that resulted from the digestion by HindIII, a particle of excluded volume proportional to size of the restriction fragment (in base pairs) was assigned (Fig. 1e). For the α-globin domain, a total of 70 particles restrained by 1049 restraints were set (Fig. 1d). The restraints included 235 harmonic, 709 lower-bound harmonic, and 105 upper-bound harmonic. A total of 50,000 independent simulations were performed with IMP in a cluster of 100 CPUs for about a week of calculations. Each run returned a single model with the lowest IMP objective function of that run. We then selected and structurally clustered from the set of 50,000 models the 10,000 models with the lowest IMP objective function (that is, the models that best satisfied the input restraints), which resulted in 393 clusters, with the top 10 largest clusters accounting for 26% of the selected models (Fig. 3a). The diversity in the solutions could be related to variable karyotypes and gene expression in individual cells in this cancer-derived cell line. The resulting clusters of models were indirectly validated by: (i) mapping local chromatin structural features from the ENCODE consortium [36] (e.g., DNase I sensitivity) and (ii) performing FISH experiments, which validated the models' overall size and shape [18].

5. Conclusions

We have shown that the integration of 5C experiments with IMP can be used to determine the 3D conformation of chromosomal domains. Our method can be used to determine the structure of genomic domains and genomes at unprecedented resolution, giving important insights into genomic long-range interactions. This information can be useful in annotating genes and their regulatory elements along the genome linear sequence, as well as in providing insights into functional relationships between genes and their distant regulatory elements, as seen in the α- globin domain test case [18]. Indeed, we have previously shown that the generated 3D structures of the α-globin domain suggests a model for higher-order chromatin folding based on the formation of chromatin globules, where actively transcribed genes clustered. The discovery of chromatin globules shows that our models can reveal novel higher-order features of chromosome architecture and could further help the large-scale efforts in the annotation of genes and regulatory elements along the linear genome. Moreover, using Fluorescent *In Situ* Hybridization (FISH) imaging we validated the proposed globular structure of the α-globin domain. Such experiments showed that for K562 cell lines we observed two globules, which were not observed in the GM12878 cell line that has the α-globin genes silent [18].

Acknowledgments

We thank the Dekker group for their support during the development of our approach. We also thank the IMP community (http://www.integrativemodeling.org) and the Chimera developers (http://www.cgl.ucsf.edu/chimera). Financial support from the Spanish Ministerio de Ciencia e Innovación (BFU2010-19310/BMC) and the Human Frountiers Science Program (RGP0044/2011) is also acknowledged. This article was partially based on the authors' previous work [18,37].

References

[1] T. Takizawa, K.J. Meaburn, T. Misteli, Cell 135 (2008) 9–13.
[2] J. Dekker, K. Rippe, M. Dekker, N. Kleckner, Science 295 (2002) 1306–1311.
[3] J. Mateos-Langerak, M. Bohn, W. de Leeuw, O. Giromus, E.M. Manders, P.J. Verschure, M.H. Indemans, H.J. Gierman, D.W. Heermann, R. van Driel, S. Goetze, Proc. Natl. Acad. Sci. USA 106 (2009) 3812–3817.
[4] G. Wedemann, J. Langowski, Biophys. J. 82 (2002) 2847–2859.
[5] J. Dekker, J. Biol. Chem. 283 (2008) 34532–34540.
[6] M. Wachsmuth, M. Caudron-Herger, K. Rippe, Biochim. Biophys. Acta 1783 (2008) 2061–2079.
[7] J. Langowski, D.W. Heermann, Semin. Cell Dev. Biol. 18 (2007) 659–667.
[8] A. Rosa, R. Everaers, PLoS Comput. Biol. 4 (2008) e1000153.
[9] J. Rouquette, C. Cremer, T. Cremer, S. Fakan, Int. Rev. Cell Mol. Biol. 282 (2010) 1–90.
[10] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, J. Dekker, Science 326 (2009) 289–293.
[11] J. Dostie, Y. Zhan, J. Dekker, Curr. Protoc. Mol. Biol. (2007) (Chapter 21, Unit 21.14).
[12] Z. Zhao, G. Tavoosidana, M. Sjolinder, A. Gondor, P. Mariano, S. Wang, C. Kanduri, M. Lezcano, K.S. Sandhu, U. Singh, V. Pant, V. Tiwari, S. Kurukuti, R. Ohlsson, Nat. Genet. 38 (2006) 1341–1347.
[13] M. Simonis, P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, W. de Laat, Nat. Genet. 38 (2006) 1348–1354.
[14] J. Dostie, J. Dekker, Nat. Protoc. 2 (2007) 988–1002.
[15] J. Dostie, T.A. Richmond, R.A. Arnaout, R.R. Selzer, W.L. Lee, T.A. Honan, E.D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R.D. Green, J. Dekker, Genome Res. 16 (2006) 1299–1309.
[16] Z. Duan, M. Andronescu, K. Schutz, S. McIlwain, Y.J. Kim, C. Lee, J. Shendure, S. Fields, C.A. Blau, W.S. Noble, Nature 465 (2010) 363.
[17] M.A. Umbarger, E. Toro, M.A. Wright, G.J. Porreca, D. Bau, S.H. Hong, M.J. Fero, L.J. Zhu, M.A. Marti-Renom, H.H. McAdams, L. Shapiro, J. Dekker, G.M. Church, Mol. Cell 44 (2011) 252–264.
[18] D. Bau, A. Sanyal, B.R. Lajoie, E. Capriotti, M. Byron, J.B. Lawrence, J. Dekker, M.A. Marti-Renom, Nat. Struct. Mol. Biol. 18 (2011) 107–114.
[19] J. Fraser, M. Rousseau, S. Shenker, M.A. Ferraiuolo, Y. Hayashizaki, M. Blanchette, J. Dostie, Genome Biol. 10 (2009) R37.
[20] R. Kalhor, H. Tjong, N. Jayathilaka, F. Alber, L. Chen, Nat. Biotechnol. 30 (2011) 90–98.
[21] Daniel Russel, Keren Lasker, Ben Webb, Javier Velázquez-Muriel, Elina Tjioe, Dina Schneidman-Duhovny, Bret Peterson, A. Sali, PLoS Biol. 10 (2012) e1001244.
[22] B. Webb, K. Lasker, D. Schneidman-Duhovny, E. Tjioe, J. Phillips, S.J. Kim, J. Velazquez-Muriel, D. Russel, A. Sali, Methods Mol. Biol. 781 (2011) 377–397.
[23] F. Alber, S. Dokudovskaya, L.M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B.T. Chait, M.P. Rout, A. Sali, Nature 450 (2007) 683–694.
[24] J. Dekker, Nat. Methods 3 (2006) 17–21.
[25] S.E. Gerchman, V. Ramakrishnan, Proc. Natl. Acad. Sci. USA 84 (1987) 7802–7806.
[26] A. Rosa, N.B. Becker, R. Everaers, Biophys. J. 98 (2010) 2410–2419.
[27] T.C. Voss, G.L. Hager, Biochim. Biophys. Acta 1783 (2008) 2044–2051.
[28] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Science 220 (1983) 671–680.
[29] N. Metropolis, S. Ulam, J. Am. Stat. Assoc. 44 (1949) 335–341.
[30] Y. Zhang, J. Skolnick, Proteins 57 (2004) 702–710.
[31] A.J. Enright, S. Van Dongen, C.A. Ouzounis, Nucleic Acids Res. 30 (2002) 1575–1584.
[32] J.B. Lawrence, R.H. Singer, J.A. McNeil, Science 249 (1990) 928–932.
[33] J.R. Hughes, J.F. Cheng, N. Ventress, S. Prabhakar, K. Clark, E. Anguita, M. De Gobbi, P. de Jong, E. Rubin, D.R. Higgs, Proc. Natl. Acad. Sci. USA 102 (2005) 9830–9835.
[34] D.R. Higgs, D. Vernimmen, J. Hughes, R. Gibbons, Annu. Rev. Genomics Hum. Genet. 8 (2007) 299–325.
[35] D.R. Higgs, W.G. Wood, Curr. Opin. Hematol. 15 (2008) 176–183.
[36] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, E.H. Margulies, Z. Weng, M. Snyder, E.T. Dermitzakis, R.E. Thurman, M.S. Kuehn, C.M. Taylor, S. Neph, C.M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J.A. Greenbaum, R.M. Andrews, P. Flicek, P.J. Boyle, H. Cao, N.P. Carter, G.K. Clelland, S. Davis, N. Day, P. Dhami, S.C. Dillon, M.O. Dorschner, H. Fiegler, P.G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K.D. James, B.E. Johnson, E.M. Johnson, T.T. Frum, E.R. Rosenzweig, N. Karnani, K. Lee, G.C. Lefebvre, P.A. Navas, F. Neri, S.C. Parker, P.J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F.S. Collins, J. Dekker, J.D. Lieb, T.D. Tullius, G.E. Crawford, S. Sunyaev, W.S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I.L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H.A. Hirsch, E.A. Sekinger, J. Lagarde, J.F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J.S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M.C. Dickson, D.J. Thomas, M.T. Weirauch, J. Gilbert, et al., Nature 447 (2007) 799–816.
[37] D. Bau, M.A. Marti-Renom, Chromosome Res. 19 (2011) 25–35.
[38] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J. Comput. Chem. 25 (2004) 1605–1612.