The 3D folding of the α-globin gene domain reveals formation of chromatin globules

Davide Baù[1], Amartya Sanyal[2], Bryan Lajoie[2], Emidio Capriotti[1], Meg Byron[3], Jeanne B. Lawrence[3], Job Dekker[2] and Marc A. Marti-Renom[1]

1. Structural Genomics Unit, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, 46012 Valencia, Spain.
2. Department of Biochemistry and Molecular Pharmacology, Program in Gene Function and Expression, University of Massachusetts Medical School 01605-2324 Worcester MA, USA.
3. Department of Cell Biology, University of Massachusetts Medical School 01605-2324 Worcester MA, USA.

Abstract

We developed a general approach that combines comprehensive chromatin interaction mapping using 5C (http://my5c.umassmed.edu)[1] with the Integrated Modeling Platform (IMP, http://www.integrativemodeling.org)[2] to generate high-resolution three-dimensional (3D) models of chromatin at the Mb scale. We applied this approach to analyze a 500 Kb gene dense domain (ENm008) on human chromosome 16 that includes the α-globin locus. We obtained 3D models of this domain in cells that express the α-globin locus (K562) as well as in lymphoblastoid cells that do not (GM12878). The models accurately reproduce the known looping interactions between the α-globin genes and their distal regulatory elements. Further, we find that the domain folds into a single globular conformation in GM12878 cells, whereas two globules are formed in K562 cells. The central cores of these globules are enriched for actively transcribed genes, whereas non-transcribed chromatin is more peripheral. We propose that globule formation represents a higher order folding state that may be related to clustering of actively transcribed genes around shared transcription machineries as observed by microscopy.

Methods

A comprehensive interaction map of the α-globin locus was obtained by performing 5C analysis of the ENm008 ~500 Kb ENCODE region. 5C experiments were performed in GM12878 cells where α-globin genes are not expressed and in K562 cells where α-globin genes are expressed and long-range interactions are expected to occur between the active genes and their upstream hypersensitive sites (HS). The 5C interaction map detected all previously known long-range looping interactions (Figure 1). 5C data was then translated into spatial restraints between points representing the restriction fragments within the ~500 Kb domain. The 70 fragments representing the studied region were restrained with a total of 1,520 and 1,049 harmonic oscillators for GM12878 and K562 cells, respectively (Figure 2). Forces applied to the defined restraints were also set proportional to the absolute value of the 5C Z-score observed between a pair of fragments. Once the α-globin locus was represented by a set of fragments and the spatial restraints between them, IMP expressed the problem of determining its structure as an optimization problem, where the final models try to satisfy most of the imposed spatial restraints between the fragments. GM12878 models clustered in a total of 4 different conformations while K562 models clustered around a more variable set of solutions with 393 different structure clusters including 10 large clusters with more than 150 solutions each and 194 clusters with less than 10 solutions each (Figure 3).

Results

We assessed the reliability of our models by determining whether they reproduce known experimental observations of chromatin looping within the ENm008 region. For example, the interaction between the restriction fragment containing the α-globin enhancer (HS40) and α-globin genes was present in all of the models in the K562 selected cluster. Conversely, the restriction fragments containing most of the α-globin genes was in close spatial proximity to all of the fragments containing the upstream DNAseI hypersensitive, which regulate their expression. Also, we find that CTCF sites are abundant in the compact region around the α-goblin genes. Additionally, it has been proposed that the α-globin genes and the group of surrounding house keeping genes associate with a common transcription factory[3]. Interestingly, we note that this entire region appear to be wrapped around a cavity with an average diameter of ~110 nm, which closely corresponds to the estimated size of a transcription factory (i.e., ~87 nm)[4]. Models were independently validated by Fluorescence *In Situ* Hybridization (FISH), which confirmed that in GM12878 cells the ENm008 region forms a single chromatin globule whereas in K562 cells the locus forms two chromatin globules (Figure 4).

Conclusion

We have demonstrated that 5C experiments combined with IMP can be used to determine the range of 3D conformations of defined chromosomal domains. By providing insights into long-range relationships between distant genomic loci our approach has the potential to further leverage large-scale efforts by the genomic community that aim to annotate genes and their regulatory elements along the linear genome by being able to determine the relative spatial arrangements of these elements. Our analysis of the α-globin locus illustrates how spatial models based on *in vivo* chromatin interaction data can provide insights into functional relationships between genes and their distant regulatory elements. Further, the identification of chromatin globules indicates how these models can point to the presence of novel higher-order features of chromosome architecture.

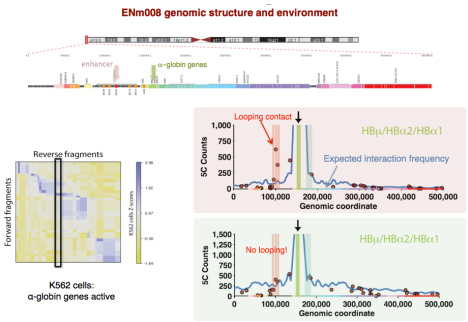

Figure 1. 5C experimental data

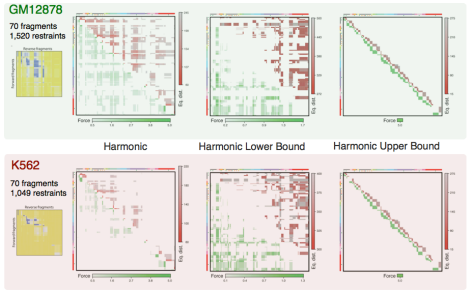

Figure 2. IMP restraints

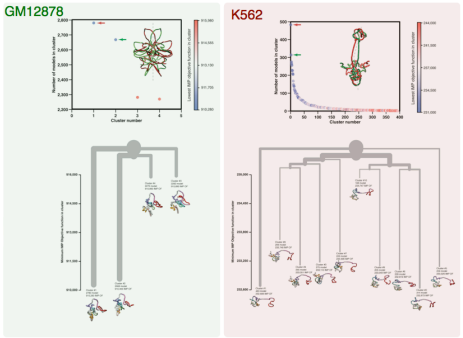

Figure 3. IMP computational modeling and analysis

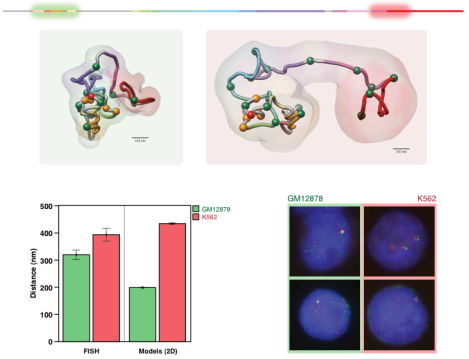

Figure 4. α-globin 3D structure models for K562 and GM12878 cell lines

References

1. J. Dostie et al., Genome Res 16, 1299 (2006).
2. F. Alber et al., Nature 450, 683 (2007)
3. G. L. Zhou et al., Mol Cell Biol 26, 5096 (2006).
4. C. H. Eskiw, A. Rapp, D. R. Carter, P. R. Cook, J Cell Sci 121, 1999 (2008).