From 5C to 3D: a hybrid method for determining the structure of genomes

Davide Baù[1], Amartya Sanyal[2], Bryan Lajoie[2], Emidio Capriotti[1], 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.

Aim A new challenge for genome studies is to establish functional relationships between genes and distant regulatory elements. Loci located far apart along the chromosome can engage in long-range looping interactions. Thus, the spatial arrangement of chromatin can provide novel insights into the long-range relationships between distant loci. We aim at developing determining the three-dimensional (3D) structure of genomic domains by integrating diverse experimental analysis within a computational framework. To do so, we have recently adapted the Integrative Modeling Platform (IMP, http://www.integrativemodeling.org)[1] to model the 3D structure of the human α-globin domain using 5C analysis (http://my5c.umassmed.edu)[2].

Methods A comprehensive interaction map of the α-globin locus (Fig. 1A) was obtained by performing 5C analysis of the ENm008 ~500 Kb ENCODE region. 5C experiments were performed in α-globin expressing K562 cells where long-range interactions are expected to occur between the α-globin genes and the upstream hypersensitive (HS) elements (Fig. 1B). The 5C interaction map detected all previously known long-range looping interactions (Fig. 1B,C,D). 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 1,049 restraints (Fig. 1E). 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. Our experiment resulted in 10,000 3D models (Fig. 2A,B), which were grouped in a total of 393 different structure clusters including ten large clusters with more than 150 solutions each and 194 clusters with less than 10 solutions each (Fig. 2C).

Results We assessed the reliability of our models (Fig. 3A) 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 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-gobin 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 (Fig. 3B), which closely corresponds to the estimated size of a transcription factory (i.e., ~87 nm)[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, (*e.g.*, between genes and their long-range regulatory elements), 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.

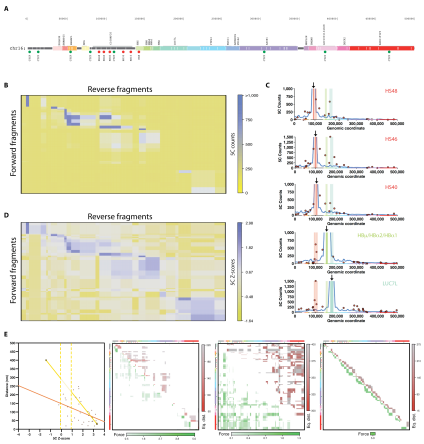

Figure 1. 5C experimental data analysis

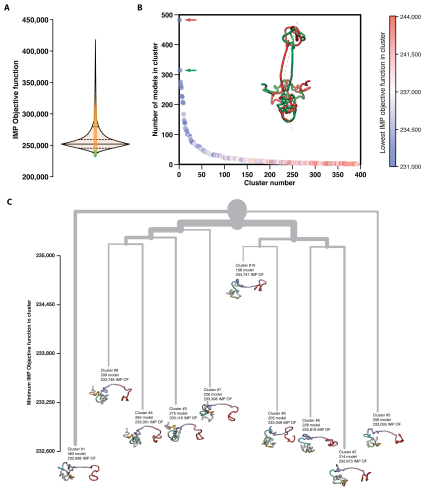

Figure 2. IMP computational modeling and analysis

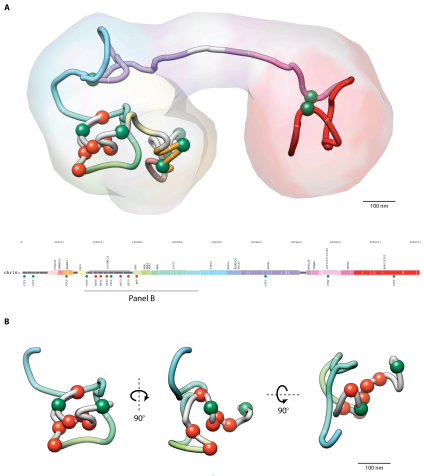

Figure 3. α-globin 3D structure models.

References

1. F. Alber et al., Nature 450, 683 (2007)
2. J. Dostie et al., Genome Res 16, 1299 (2006).
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).