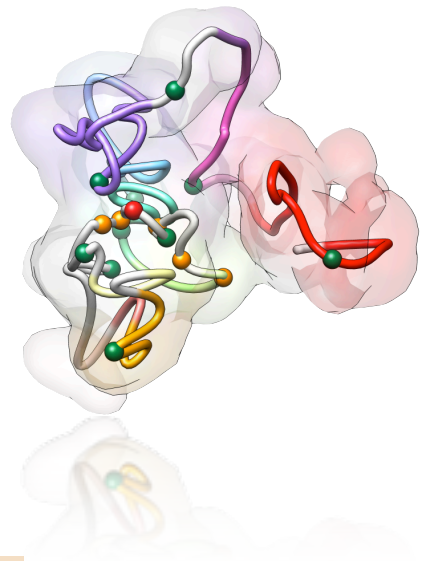
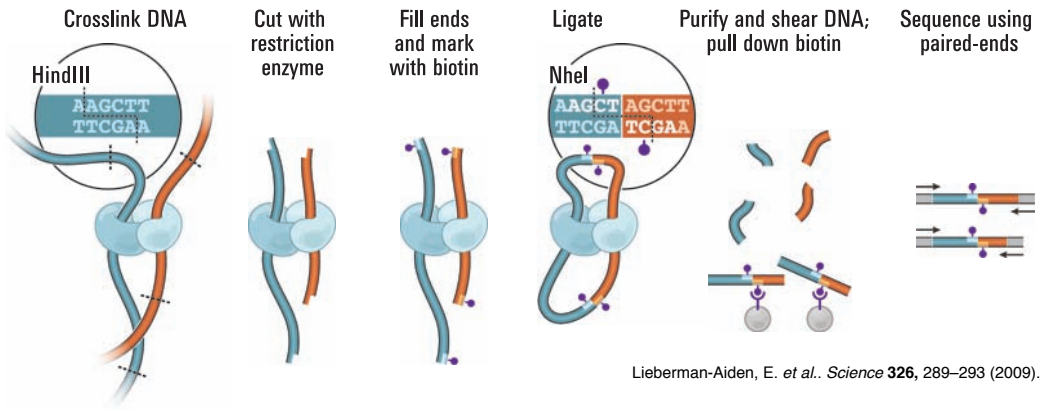
TADBIT, A PIPELINE FOR THE 3D MODELING OF GENOMES AND GENOMIC DOMAINS USING 3C-BASED INTERACTION MATRICES

Davide Baù[1,2], François Serra[1,2], Guillaume Filion[2] and Marc A. Marti-Renom[1,2]

[1]Genome Biology Group. National Center for Genomic Analysis (CNAG), Barcelona, Spain. [2]Gene Regulation, Stem Cells and Cancer Program, Centre de Regulació Genòmica (CRG), Barcelona, Spain

***Introduction**:* The sequence of a genome alone does not carry the information needed for understanding how genomic processes are carried out in the cell nucleus; to achieve this, the knowledge of its three-dimensional (3D) architecture is necessary. Advances in genomic technologies and the development of new methods, such those based on Chromosome Conformation Capture (3C) [1], have allowed getting insights at unprecedented resolution into how genomes are organized. Recently, it has been shown that chromatin is organized in Topologically Associating Domains (TADs), large interacting domains that appear to be conserved among different cell types. Here we describe TADBit, a pipeline for the 3D modeling of genomic domains; TADBit is python library that extends the Integrative Modeling Platform (IMP) [2] to determine the 3D architecture of genomic domains and entire genomes using chromosome conformation capture data [3].





Lieberman-Aiden, E. *et al.*. *Science* **326,** 289–293 (2009).

3C-based interaction matrices

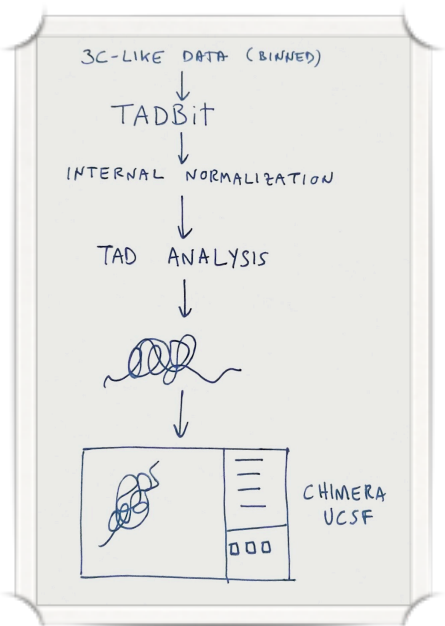
TADBit has been designed to work with binned interaction matrices from 3C-based experiments such as Hi-C data [4]. The Hi-C technique allows to investigate the proximity of loci located on the same or different chromosomes. Hi-C experimental data result in interaction counts between pairs of loci from the genomic region of interest (*i.e.*, the quantitative determination of the number of times each specific experimental ligation product is sequenced).

Hi-C matrices can be segmented into TADs

TADBit normalizes the Hi-C raw interactions by correcting local biases derived by G+C content, availability of the restriction enzyme site and repeat coverage, which could affect the read count of a bin. The normalization procedure is based on the hiclib (https://bitbucket.org/mirnylab/hiclib) software [5]. Once the data have been normalized, TADBit analyzes the contact distribution along the genome and subsequently segments it into its constitutive TADs. The user will select TADs from the list generated by TADBit; along with the normalized interaction data for the selected TADs, TADBit will generate a heat-map plot of the region for a quick visual inspection of the input data.

Modeling the 3D structure of selected TADs

Each modeled TAD is represented by a set of particles, one per experimental (Hi-C) bin. Each particle has a radius proportional to the number of bases in the bin. TADBit determines the 3D structure of a TAD by translating the interaction matrix into spatial restraints between each particle and by building an ensemble of structures that satisfies the imposed restraints. Additionally, TADBit implement many diverse sets of functions for analyzing the resulting 3D models.



Data functions:
- Data cleaning
- Data normalization
- TAD identification
- TAD comparison
- TAD clustering

3D modeling:
- IMP optimization
- 3D modeling

3D analysis:
- Clustering of models
- Density plots
- Distance/angle plots
- Hi-C correlation
- etc...

Summary: TADBit, a new computational python library, has been developed to build 3D models of genomic domains and genomes using 3C-based interaction matrices. Additionally, TADBit has several functions that allow the user to easily select parts of the interaction matrix (such as TADs) and to model and analyze their structures. Please, write to mmarti@pcb.ub.cat if your are interested in getting TADBit.

Bibliography

1. *Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. Science (New York, NY) 295, 1306–1311 (2002).*
2. *Russel, D., et al. (2012). Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. PLoS Biology, 10(1), e1001244*
3. *Baù, D. & Marti-Renom, M. A. Genome structure determination via 3C-based data integration by the Integrative Modeling Platform. Methods 58, 300–306 (2012).*
4. *Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (New York, NY) 326, 289–293 (2009).*
5. *Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Meth (2012). doi:10.1038/nmeth.2148*

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG Centre de Regulació Genòmica

HFSP

>> more information at http://marciuslab.org <<