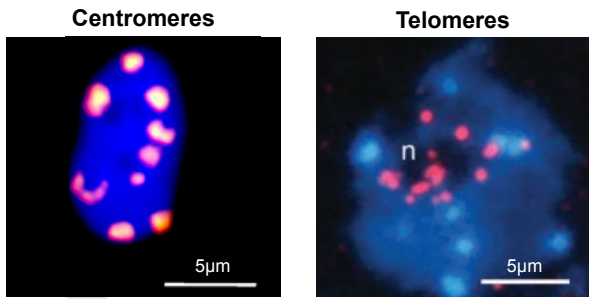
KNOWLEDGE-BASED MODELING OF ARABIDOPSIS THALIANA GENOME

Marco Di Stefano[1], Davide Baù[1], François Serra[1] and Marc A. Marti-Renom[1,2]

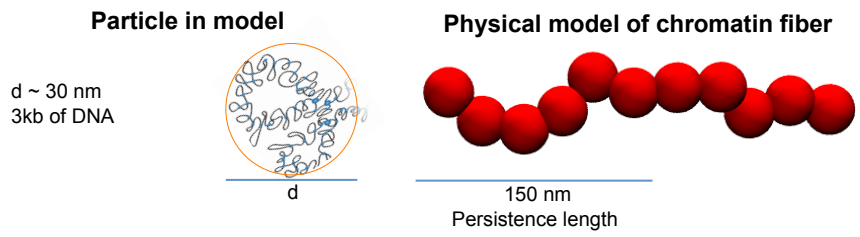
[1]Structural Genomics group. Centre Nacional d'Anàlisi Genòmica - Centre de Regulació Genòmica (CNAG-CRG), Barcelona, Spain. [2] Institució de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

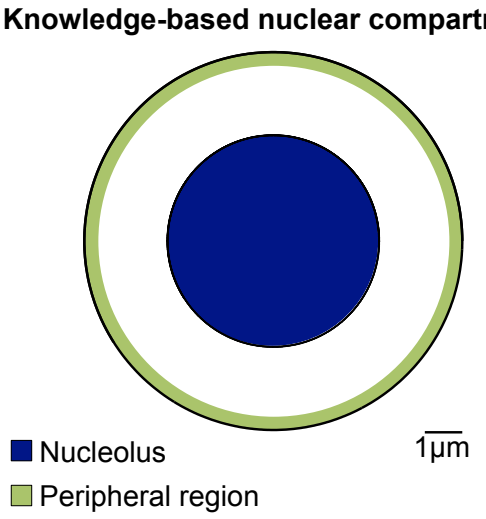
***Introduction**:* The spatial organisation of the *Arabidopsis thaliana* genome has been studied in details using imaging techniques [1,2]. These approaches unveiled distinctive features in the *A. thaliana* large-scale genome organisation, such as the preferential positioning of the centromeres at the nuclear periphery [1] and the proximity of the telomeres to the nucleolus [2]. Here, we combine this information with physical models of chromatin and steered molecular dynamics (steered MD) simulations [3,4] to address the role of the large-scale spatial restraints in the 3D genome arrangement, at the kilo base resolution. The obtained 3D models are validated against recently published Hi-C data [5-7] using TADbit, a bioinformatics tool developed by our group [8]. This strategy will provide useful biological insights into the structural arrangement of the *A. thaliana* genome highlighting the key features of its organisation.



Chromatin model

The chromatin of *A. thaliana* is represented as a chain of connected particles characterised by excluded volume and bending rigidity interactions. The parameters of the model are set to describe the nominal physical properties of the 30nm chromatin fiber [3,4]. Experimentally obtained restraint [1,2] are used to guide the possible arrangements of the polymer during the steered MD simulations.

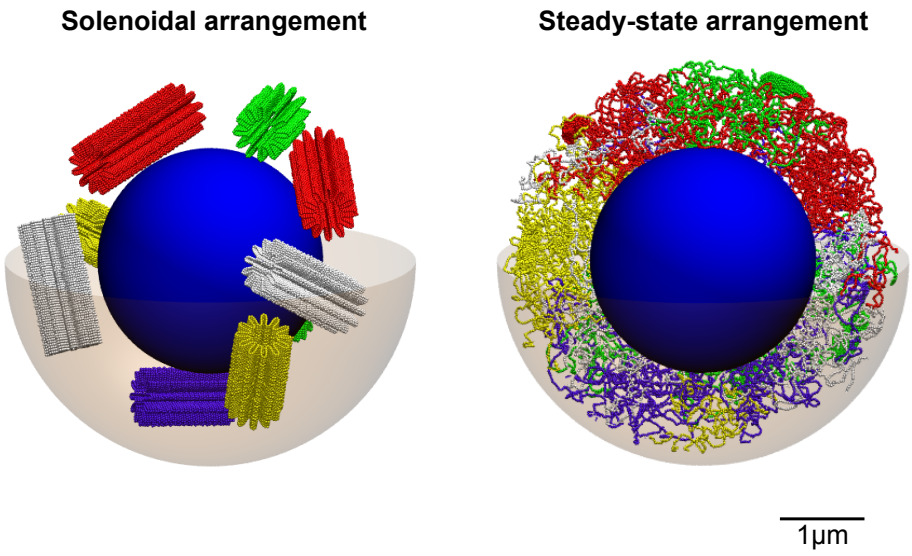




3D genome model

Each chromosome chain is initially arranged in a solenoidal structure forming rosettes around the principal axis [3,4]. The resulting elongated structures are randomly positioned and orientated within the nuclear space (represented as a sphere), while the nucleolus is maintained void of particles.

During the steered MD simulation a set of additional restraints (harmonic potentials) [1,2] are used to guide the 3D genome arrangement. These restraints are applied to tether the telomeres to the nucleolar periphery and to constrain the motion of the centromeres within the peripheral region. Independent systems of chromosome chains are simulated up to a steady-state arrangement, in which all the spatial restraints are simultaneously satisfied.



Model validation using Hi-C experimental data

The obtained set of conformations are validated against publicly available Hi-C data [5-7]. Hi-C data are analysed with TADbit [8], a bioinformatics tool designed to study 3C-like data. TADbit covers all the steps of the Hi-C analysis from the mapping of the sequences onto the reference genome to the normalisation of the raw data by correcting for local biases. Here, we will develop and test new features in TADbit such as the possibility to compare matrices collected in different experiments.

Summary: The aim of this ongoing project is to study the role of general polymer physics and large-scale spatial restraints (from imaging experiments) in the 3D arrangement of *A. thaliana* genome. The final models are validated using recent Hi-C data and bioinformatics tools. This approach will provide a characterisation of the interplay between the 3D genome organisation and functionally related features of chromatin in *A. thaliana*.

Bibliography

1. *Fransz, P. et al. Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. PNAS U S A. 99(22): 14584-9 (2002).*
2. *Schubert, V. et al. Interphase chromatin organisation in Arabidopsis nuclei: constraints versus randomness. Chromosoma 121:369-387 (2012).*
3. *Rosa, A. et al. Looping probabilities in model interphase chromosomes. Biophys J 98(11):2410-9 (2010).*
4. *Di Stefano, M. et al. Colocalization of Coregulated Genes: A Steered Molecular Dynamics Study of Human Chromosome 19. PLoS Comput Biol. 9(3):e1003019 (2013).*
5. *Grob, S. et al. Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. Mol Cell. 55(5):678-93 (2014).*
6. *Feng, S. et al. Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. Mol Cell. 55(5):694-707 (2014).*
7. *Wang, C. et al. Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome Res. 25:246-56 (2015).*
8. *D. Baù and M.A. Marti-Renom. Structure determination of genomic domains by satisfaction of spatial restraints. Chromosome Res. 19(1):25-35 (2011).*

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

CRG Centre for Genomic Regulation

erc European Research Council
Established by the European Commission

European Commission

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