



Multiscale Complex Genomics



**Project Acronym:** MuG

**Project title:** Multi-Scale Complex Genomics (MuG)

**Call:** H2020-EINFRA-2015-1

**Topic:** EINFRA-9-2015

**Project Number:** 676556

**Project Coordinator:** Institute for Research in Biomedicine (IRB Barcelona)

**Project start date:** 1/11/2015

**Duration:** 36 months

## **Milestone 21: Production of the first 3D/4D models of protein-DNA complexes.**

**Lead beneficiary:** University of Nottingham

**Dissemination level:** PUBLIC

**Due date:** 01/05/2016

**Actual submission date:** 20/05/2016

Copyright© 2015-2018 The partners of the MuG Consortium



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.

## Document History

Version	Contributor(s)	Partner	Date	Comments
0.1	Marco Pasi	UNOT	26/04/16	First draft
0.2	Marco Pasi	UNOT	16/05/16	Final draft
0.3				

An important objective of the MuG project is to provide the infrastructure to organise and store 3D/4D models of protein-DNA complexes, as well as useful end-user interfaces to quickly and effectively access this data for researchers in the field of multiscale genomics. After careful consideration of available data repositories, BIGNASim (Hospital et al., 2015) was chosen to perform these tasks in the MuG project, as it complies with the requirements set out in the Data Management Plan for this type of data (for more details please refer to <https://3.basecamp.com/3126297/buckets/97795/uploads/122576165>).

A MuG-specific section of BIGNASim was created to host data that is produced within the project, and that will respect the policies for data access, interoperability and public release defined in the Data Management Plan. As a first milestone in Work Package 6, we have made available the first 3D/4D models of protein-DNA complexes through the MuG section of the BIGNASim repository. The aim of this work has been to technically evaluate the full procedure of making these models available to the public, in order to identify as early as possible issues that may require further action.

This initial data, in the form of a molecular dynamics (MD) simulation of the telomere repeat binding factor TRF1 (Garton et al., 2013) bound to its target sequence, is publicly available as raw data to download from the MuG section of BIGNASim, which will be fully integrated in the MuG portal at a later stage (<http://mmb.irbbarcelona.org/BigNASimMuG>). More MD simulations will be made available in the same way as the project progresses: either run as part of Pilot Project 7.3, but also uploaded to the MuG repository by researchers who want to make their simulation data available. MuG users will have access to this large body of simulation data, and in cases where this may be required, to the rest of the data that is present in the BIGNASim repository but not directly related to multiscale genomics.

In addition to their availability in raw format, a wide array of analyses have been run on the MD simulation, using the tools available in BIGNASim, to improve the accessibility of the simulation data to researchers outside the field of molecular simulations. The results of these analyses have been stored in the MuG repository, and are accessible through the rich web interface provided by BIGNASim. A complete offer of tools to analyse protein-DNA complex simulations will be integrated in the VRE in the next six months, and detailed as part of Deliverable 6.2.

## References

Hospital A, Andrio P, Cugnasco C, Codo L, Becerra Y, Dans PD, et al. BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.* 2016;44: D272–8.

Garton, M., & Laughton, C. A Comprehensive Model for the Recognition of Human Telomeres by TRF1. *J Mol Biol.* 2013;425(16): 2910–2921.