# Abstract

Development of chromosome conformation capture (3C) and its derivative techniques such as Hi-C and ChIA-PET allow to capture the details of the genome organization on multiple scales and led to many new insights into the mechanisms of genome functioning. It is already well established that genome organization is closely linked to genomic processes such as gene expression regulation, cell development and cell differentiation.

Moreover, a number of diseases, including multiple types of cancers, are directly caused by disruptions of this organization induced by mutations targeting structural genomic elements. This adds an additional dimension to genomic analyses, as even subtle mutations, for instance single nucleotide polymorphisms located in non-coding regions may significantly impact genes located in distant parts of the genome. To fully understand how genome functions, both in health and disease, a thorough knowledge of the genomic structure and its dynamics is thus required.

Data from 3C experiments is usually presented using simple 1D and 2D representations such as interaction networks and contact maps. While these representations are useful, 3D models provide much more comprehensive view for studying genomic features such as chromatin accessibility or compaction. However, due to biological, experimental and computational limitations, it is not an easy task to create high-resolution, biologically feasible 3D models.

A number of computational approaches for 3D genome modeling were developed since the advent of the 3C methods. These approaches, however, are often based solely on the contact frequency data from the experiments capturing non-specific interactions (such as Hi-C) without incorporating current knowledge about genome organization in the modeling, and thus they may not be able to adequately reflect the underlying biological structures. Moreover, due to computational reasons the simulation size is often limited, which means that these methods can be applied only to low-resolution data or to modeling short genomic regions.

This work presents 3D-GNOME (**3**-**D**imensional **G**e**no**me **M**odeling **E**ngine), a novel multi-scale algorithm for modeling 3D genome organization which is able to generate high-resolution whole genome models in a reasonable time. The most prominent feature of the algorithm - and one that probably makes it most distinct from the existing methods - is a very close coupling to the biological model of genome topology. 3D-GNOME works in a hierarchical fashion, where every hierarchy level corresponds to a specific scale of the genome organization (chromosome territories, topological domains, and chromatin loops) and beads used for modeling directly correspond to biological structures, as opposed to majority of the existing methods, where beads correspond to arbitrary fragments of the genome without a biological meaning. Thus, the algorithm generates models that can be viewed and analyzed at different biological scales, providing a comprehensive view of the genome topology.

The algorithm was developed specifically for ChIA-PET, an advanced 3C technique that allows to capture chromatin interactions mediated by a specific protein factor, in principle, however, it can also be used with other types of data such as 5C or Hi-C. To present the capabilities of the algorithm a number of models at different resolutions were created and analyzed using a combined CTCF and RNAPII ChIA-PET dataset from a GM12878 cell line. Given the important role these factors play in genome (CTCF performs a major role in organizing chromatin, whereas RNAPII is involved in transcription of many genes) the generated structures provide a comprehensive view of genome structural and functional features.

The algorithm was developed both as a stand-alone software package, which allows to easily include the algorithm in computational pipelines or to run a large number of simulations using scripting, and as a user-friendly web-server, which facilitates the use of the algorithm for non-technical users.