Structural bioinformatics

3DPolyS-LE: an accessible simulation framework to model the interplay between chromatin and loop extrusion

Todor Gitchev ◉ [1,†], **Gabriel Zala**[1], **Peter Meister** ◉ [1,*] **and Daniel Jost** ◉ [2,*]

[1]Cell Fate and Nuclear Organization, Institute of Cell Biology, University of Bern, Bern 3012, Switzerland and [2]Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR5239, Inserm U1293, Université Claude Bernard Lyon 1, Lyon 69007, France

*To whom correspondence should be addressed.

†Present address: Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland

Abstract

Summary: Recent studies suggest that the loop extrusion activity of Structural Maintenance of Chromosomes complexes is central to proper organization of genomes *in vivo*. Polymer physics-based modeling of chromosome structure has been instrumental to assess which structures such extrusion can create. Only few laboratories however have the technical and computational expertise to create *in silico* models combining dynamic features of chromatin and loop extruders. Here, we present 3DPolyS-LE, a self-contained, easy to use modeling and simulation framework allowing non-specialists to ask how specific properties of loop extruders and boundary elements impact on 3D chromosome structure. 3DPolyS-LE also provides algorithms to compare predictions with experimental Hi-C data.

Availability and implementation: Software available at https://gitlab.com/togop/3DPolyS-LE; implemented in Python and Fortran 2003 and supported on any Unix-based operating system (Linux and Mac OS).

Contact: peter.meister@unibe.ch or daniel.jost@ens-lyon.fr

Supplementary information: Supplementary information are available at *Bioinformatics* online.

1 Introduction

Genes are regulated at many levels, from local transcription factor binding to the megabase-range contacts between enhancers and promoters. Recent findings have highlighted the function of the chromosome 3D organization in the latter regulation, as the genome is partitioned into consecutive regions of enhanced compaction, the so-called 'topologically associated domains' (TADs), where promoters and enhancers colocalize (Dixon *et al.*, 2012; Nora *et al.*, 2012). At this scale, genome folding is mostly a consequence of the interplay between loop extrusion factors of the Structural Maintenance of Chromosome (SMC) complexes family and oriented boundary elements bound by proteins that limit loop extrusion [reviewed in Mirny and Dekker (2022)]. In particular, by comparing results from *in silico* models and *in vivo* Hi-C data, polymer simulations proved very useful to understand the TAD structure of chromosomes, as well as to suggest and test hypotheses on the function of boundary elements or loop extrusion factors (Brandão *et al.*, 2021; Fudenberg *et al.*, 2016; Goloborodko *et al.*, 2016; Nuebler *et al.*, 2018; Rao *et al.*, 2017; Schwarzer *et al.*, 2017). As the development of such simulations is technically difficult and thus generally not accessible to biologists aiming to (in)validate a mechanistic

hypothesis on TAD formation for their system of interest, we provide an open-access, user-friendly, generic modeling framework for physics-based polymer simulations of loop extrusion (3DPolyS-LE), wrapped as a Python package, allowing users to run simulations, varying parameters on boundary elements and loop extruders properties and assess the expected structures.

2 Model

3DPolyS-LE simulates the dynamics of one chromosome, modeled as a coarse-grain polymeric chain in which each monomer, of size 50 nm, contains 2 kb of chromatinized DNA. In absence of loop extrusion, the dynamics of the chain is governed by the generic properties of a homopoly-mer: chain connectivity, excluded volume and bending rigidity (Ghosh and Jost, 2018). Additionally, the polymer can be extruded by loop extruding factors (LEFs) that dynamically bind and unbind from chromatin (Fig. 1A, Supplementary Methods). Initial binding of LEFs could be at predefined loading sites or non-specifically along the chromosome. Bound LEFs are composed of two 'legs' that may translocate along the genome, creating dynamic loops between gradually more distant regions along
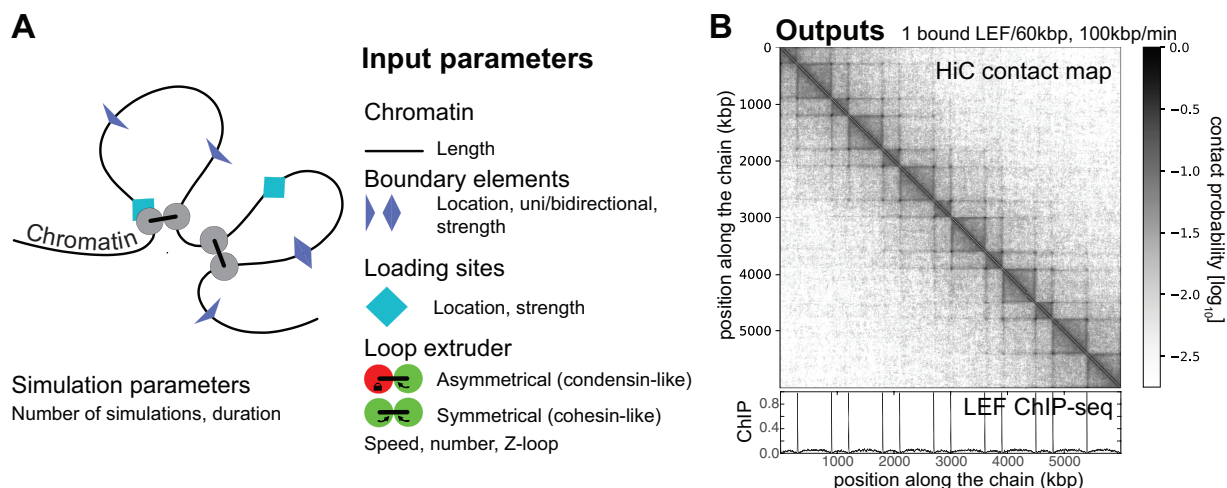
Fig. 1. Features provided by 3DPolyS-LE. (**A**) Input parameters for the simulation framework. For the polymer, its length, as well as the location and permeability of individual loading sites and boundary elements for the loop extrusion factors (LEFs) can be defined. The properties of the LEFs include their mode of extrusion (symmetrical/asymmetrical), the extrusion speed, the number of LEFs per polymer and the capacity of LEFs to cross each other (Z-loop formation). (**B**) Typical outputs of the simulations: virtual Hi-C data (top) and ChIP-Seq profile (bottom) of loop extruders

the chain. We implemented two scenarios for the leg motion: (i) symmetric extrusion with LEF legs progressing along chromatin in opposite directions at the same speed, as observed *in vitro* for cohesin (Davidson *et al.*, 2019; Kim *et al.*, 2019); (ii) asymmetric extrusion with only one translocating leg, as observed *in vitro* for condensin (Ganji *et al.*, 2018; Kong *et al.*, 2020). The motion of a LEF can be restricted by the presence of boundary elements that may stop or slow down the progression of legs depending on their directionality, and by collisions with the other extruding LEFs. We integrated two scenarios for collisions between extruding legs: (i) legs are impenetrable obstacles and they cannot move until one detaches from chromatin as usually assumed for cohesin-mediated extrusion (Fudenberg *et al.*, 2016); (ii) legs are phantom obstacles and can cross each other. This is the so-called Z-loop process recently observed *in vitro* for yeast condensin (Kim *et al.*, 2020) and *in vivo* for bacterial SMCs (Brandão *et al.*, 2021).

3 Description of the program

3.1 Inputs and outputs

3DPolyS-LE allows to modify several parameters controlling LEF properties (binding, density and velocity), to select the leg motion type (symmetric/asymmetric) and head-to-head collisions scenarios (impenetrable/phantom), as well as to position of boundary elements (with individual directionality and strength; for a detailed description of parameters and how to change them, see Supplementary Methods). Depending on the model parameters and extrusion properties, 3DPolyS-LE simulates, for a given number of independent polymers, the dynamics of the chromosome during a user-defined time period. During the simulations, snapshots of the current polymer conformations and LEF positions are stored at regularly spaced time intervals. From these snapshots, virtual Hi-C maps and ChIP-seq profiles for LEF occupancy are produced in HDF5 (with an included converter to 'cooler' format; Abdennur and Mirny, 2020) and bedGraph formats, respectively (Fig. 1B). Optionally, three-way chromatin contacts (Supplementary Fig. S3) can be extracted for downstream comparative analysis with data from GAM (Beagrie *et al.*, 2017) or multi-contact nanopore-derived 'C' technologies (Allahyar *et al.*, 2018; Deshpande *et al.*, 2022). If a reference Hi-C map is provided, 3DPolyS-LE will compute relevant metrics including a χ^2-score (Supplementary Methods) to quantitatively compare model predictions with this data. A user-guide on how to select parameter ranges is given in Supplementary Methods.

3.2 Implementation and performance

The program is organized as a Python package, requiring specific libraries for compilation and parallel processing of the core simulation module and for the downstream analysis (Supplementary Methods). The first step 'Simulations' is running simulations defined by the configuration files describing parameter values, with the possibility to use multiple cores of a High-Performance Computing (HPC) cluster over a Message Passing Interface framework. The next step called 'Analysis' processes simulation results and extracts Hi-C and ChIP-seq data. The last step 'Comparison' is the comparison to a provided reference dataset and the production of related plots. In the case of a series of simulations with different parameters (or 'grid'-simulations, Supplementary Methods), summary statistics can be visualized in a heat-map plot. All steps are run with a single command using a scheduler. The package is working on any Unix-based operating system (Linux, MacOS) and has been tested on a Slurm HPC cluster. Using a polymer equivalent to a 6 Mb chromosome (3000 beads), a 2 h (real time) simulation required roughly 20 CPU.min (AMD Epyc, 4 Gb RAM).

3.3 Examples

As an illustration of 3DPolyS-LE, we modeled loop extrusion by cohesins during interphase in mammals, by simulating an arbitrary 6 Mb-long polymer with impermeable boundaries placed along the chain (alternating between every 300 and 600 kb) with symmetric leg motion, impenetrable head-to-head collisions, random loading onto the polymer and default binding/unbinding rates estimated from *in vivo* imaging data (Supplementary Methods) (Cattoglio *et al.*, 2019; Hansen *et al.*, 2017). For a density of 1 bound LEF per 60 kb and an extruding velocity of ∼100 kb/min, we observed the formation of TADs with corner peaks (Fig. 1B). Then, to illustrate the 'grid'-simulations option, we varied the density of bound LEFs and their extruding velocity on a 8 × 4 sparse parameter grid for a total running time of ∼100 CPU.day and compared the predicted intra-TAD contact probabilities to a synthetic dataset extracted from a meta-TAD analysis of the experimental GM12878 Hi-C data for 300 kb- and 600 kb-long TADs (Supplementary Fig. S1, Supplementary Methods) (Rao *et al.*, 2014). Among this grid, the optimal parameters that minimizes the χ^2-score (Supplementary Fig. S1) are (i) a density of 1 bound LEF per 85 kb twice more than *in vivo* estimations in mESC [1 bound LEF every 186–372 kb (Cattoglio et al. 2019)] and (ii) a velocity of ∼10 kb/min in the lower range of *in vitro* estimations on naked dsDNA [30–120 kb/min (Davidson *et al.*, 2019; Golfier *et al.*, 2020; Kim *et al.*, 2019)], suggesting a slowing down of extrusion in chromatinized contexts

(Gabriele *et al.*, 2022). Other examples showing the impact of different scenarios (asymmetric leg motion, phantom collisions, loading at specific sites, boundary directionality, permeable boundaries, etc.) are given in Supplementary Figure S2.

4 Conclusion

3DPolyS-LE represents a modular framework to investigate how loop extrusion impacts chromosome folding, integrating a wide range of possible scenarios, including cohesin and condensin extrusion mode, and accounting for heterogeneities in loading rates or extrusion speeds. The model allows one to predict the effect of loop extruders on the 3D folding of specific genomic regions and test mechanistic hypotheses. Further developments will provide an integrative modeling platform combining loop extrusion and phase-separation (Ghosh and Jost, 2018), the two major mechanisms of chromosome organization (Mirny and Dekker, 2022).

Funding

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

Abdennur,N. and Mirny,L.A. (2020) Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*, **36**, 311–316.

Allahyar,A. *et al.* (2018) Enhancer hubs and loop collisions identified from Single-Allele topologies. *Nat. Genet.*, **50**, 1151–1160.

Beagrie,R.A. *et al.* (2017) Complex Multi-Enhancer contacts captured by genome architecture mapping. *Nature*, **543**, 519–524.

Brandão,H.B. *et al.* (2021) DNA-Loop-Extruding SMC complexes can traverse one another in vivo. *Nat. Struct. Mol. Biol.*, **28**, 642–651.

Cattoglio,C. *et al.* (2019) Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *eLife*, **8**, e40164.

Davidson,I.F. *et al.* (2019) DNA loop extrusion by human cohesin. *Science*, **366**, 1338–1345.

Deshpande,A.S. *et al.* Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nature Biotech*, **40**, 1488–1499.

Dixon,J.R. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.

Fudenberg,G. *et al.* (2016) Formation of chromosomal domains by loop extrusion. *Cell Rep.*, **15**, 2038–2049.

Gabriele,M. *et al.* (2022) Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, **376**, 496–501.

Ganji,M. *et al.* (2018) Real-time imaging of DNA loop extrusion by condensin. *Science*, **360**, 102–105.

Ghosh,S.K. and Jost,D. (2018) How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes. *PLoS Comp. Biol.*, **14**, e1006159.

Golfier,S. *et al.* (2020) Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. *eLife*, **9**, e53885.

Goloborodko,A. *et al.* (2016) Compaction and segregation of sister chromatids via active loop extrusion. *eLife*, **5**, e14864.

Hansen,A.S. *et al.* (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife*, **6**, e25776.

Kim,E. *et al.* (2020) DNA-Loop extruding condensin complexes can traverse one another. *Nature*, **579**, 438–442.

Kim,Y. *et al.* (2019) Human cohesin compacts DNA by loop extrusion. *Science*, **366**, 1345–1349.

Kong,M. *et al.* (2020) Human condensin I and II drive extensive ATP-dependent compaction of nucleosome-bound DNA. *Mol. Cell.*, **79**, 99–114.e9.

Mirny,L.A. and Dekker,J. (2022) Mechanisms of chromosome folding and nuclear organization: their interplay and open questions. *Cold Spring Harb. Perspect. Biol.*, **14**, a040147.

Nora,E.P. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-Inactivation Centre. *Nature*, **485**, 381–385.

Nuebler,J. *et al.* (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. USA*, **115**, E6697–E6706.

Rao,S.S.P. *et al.* (2017) Cohesin loss eliminates all loop domains. *Cell*, **171**, 305–320.e24.

Rao,S.S.P. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.

Schwarzer,W. *et al.* (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, **551**, 51–56.