Analysis of the H164N mutant M^{pro} protein in interaction with the GCP-376 ligand

Maurizio Gilioli

Contents

1 ABSTRACT

2 INTRODUCTION

2.1 Chromatin as the information center of a cell

[1] All the living organisms possess DNA, which is the main molecule through which information are passed from a generation to another. Chromatin is contained inside the nucleus in an ordered manner[1] . DNA is wrapped around histones forming nucleosomes. Throughout the report, I will call the nucleomes beads, which is a term that underlines the spherical shape of the DNA-histone complex. DNA and histones are subjected to different modifications; among those, methylation is the most important modification involving DNA. Methylation in mammals occurs in specific sites of the genome, called CpGs, where a cyto- sine is connected directly to a guanine. Methy- lations of regulatory elements have been im- plicated in determining cell identity and chro- matin structure[?,?] . CTCF is a protein conserved in eukariotes and is ubiquitous in mammalians[?] . It contains a Zync-finger which binds to DNA. The act of binding is performed in cooperation with cohesins, and causes the folding of the chromatin.

2.2 ATAC-sequencing and CTCF sequencing

ATAC-sequencing is a technology that allows for the identification of open- chromatin regions[?,?]. In order to work, it requires the addition of TN-5, a hyper-active transposase. The latter is preloaded with sequencing adapters[?] to induce a contempourary reaction of fragmentation and ligation of the pieces released, in a process called segmentation. The obtained adapted fragments are then amplified and sequenced. Once the reads are generated, a peak-calling algorithm (in our case MACS-2[?]) is used to determine which portions of the genome present ATAC peaks, and areas where there are significant enrichments of aligned reads with respect to the background. A significant enrich- ment of reads is possible only in accessible regions, which are generally also the most active ones and with available sites for transcription factors binding. CTCF data, named in chapter ..., were obtained through a classical CHIP-sequencing.

2.3 Chromatin Coarse-Graining

2.4 ChromHMM allows the sequential characterization of DNA regions

ChromHMM is a tool which helps in the annotation of genomic DNA by using epigenomic information[2] . The way it learns chromatin states signatures by using a multivariate hidden Markov model: In each genomic position, it returns the most probable chromatinic state (segments) and other useful information, such as the emission/transition parameters of the states, the abundance of the states at the TSS (Transcriptional Starting Site), at the TES (Transcriptional Ending site), and other important portions of the genome (CPG islands, exons, genes). In the case of my study, ChromHMM was used to

3 Aim of the project

The project is part of the thesis whose aim is to predict matrices of contact of chromatin through the results obtained with molecular coarse-grained simulations of 2 Mb portions of the chromatin. The scope of the report is also to gather opinions and useful feedback to improve the thesis work, which will continue for another two months.

4 METHODS

4.1 Data used for the project

The data of CTCF and ATAC for the IMR90 cell line included in the paper written by Jimin and colleagues in 2023 about the Origami simulation tool[3] were used for the project (see table 1).

Cell-Type	CTCF ChiP-seq	ATAC-seq
IMR90	ENCSR000EFI	ENCSR200OML

Table 1. Table referring to the data used for the analysis. All of them were used for the training also of the Origami[3] model.

4.2 The Model

The objective was to create a polymer model with a coarse-graining resolution of 5000 bps. The region of interest was ANPEP, it had a length of 2000000 bps and was considered in human genomes. The model creation was done by using code written by Marco Di Stefano. The total Genome length was obtained from the UCSC genome browser[?] .

4.2.1 The computation of the parameters for Coarse Graining

Both parameters for the CG model (table **??**) and the CG model were calculated. The number of bps wrapping around a bead in the FS method was considered to be 150, while instead the linker portion was considered to be of length 50 bps. The thickness of a bead (of a nucleosome) was taken as equal to 10 nm, while instead the default Kuhn length was set to 50 nms. The genome densities are imposed to be the same for the FS and the CG model.

Property	Formula	Value
nuFS (DNA content of a monomer in b.)		150+50 bps
bFS (Diameter of a bead in nm)		10 nm
lkFS (Kuhn length of the chain in FS)		50 nm
NFS (Number of monomers to represent the chromosome)		30000
NkFS (Number of Kuhn lengths of the chain)		6000
rhoFS (Genome density in bp/nm³)		0.012
rhokFS (Genome density in Kuhn lengths bp/nm³)		1.2e-05
LFS (Polymer contour length)		300000 nm
LeFS (Entanglement length of the chain in nm)		8022.22 nm
Number of monomers in a Kuhn length FS		5
lkFSnuFS_bFS (DNA content of a Kuhn length FS)		1000 bp

Table 2. Parameters calculated for the Fine Scale (FS) model

Property	Formula	Value
nuCG (DNA content of a monomer in b.)	*const.*	5000 bps
bCG (Diameter of a bead in nm)	*const.*	43.0155 nm
lkCG (Kuhn length of the chain in CG)	$\sqrt{\text{lkCGnuCG\_bCG} * \text{lkFSbFS\_nuFS}}$	290.65 nm
NCG (Number of monomers to represent the chromosome)	$\frac{\text{DNA\_content}}{\nu_{CG}} \times \text{n\_copies}$	1200
sideCG (size of the cubic simulation box)	$\frac{(N_{CG} \cdot \nu_{CG}/\rho_{FS})^{1/3}}{b_{FS}}$	18.4515
NkCG (Number of Kuhn lengths of the chain)	$\frac{N_{CG} * b_{CG}}{lk_{CG}}$	177.597
rhoCG (Genome density in bp/nm³)	*const.*	0.012
rhokCG (Genome density in Kuhn lengths bp/nm)	$\frac{\rho_{CG} \cdot b_{CG}}{\nu_{CG} \cdot \text{lk}_{CG}}$	$3.55194e-07$
LCG (Polymer contour length)	$N_{CG} * b_{CG}$	51618 nm
LeCG (Entanglement length of the chain in nm)		1379.51 nm
Number of monomers in a Kuhn length CG		6.75687
lkCGnuCG_bCG (DNA content of a Kuhn length CG)		33791

Table 3. Parameters calculated for the coarse-grained (CG) model

5 RESULTS

6 DISCUSSION AND CONCLUSIONS

7 CONCLUSIONS

8 Glossary

Bead	The complex formed by the DNA and the histone proteins
TSS	Transcriptional Starting Site
TES	Transcriptional Ending site
FS	Fine Scale
CG	Coarse Graining

References

1. Paro, P. D. R., Grossniklaus, P. D. U., Santoro, D. R. & Wutz, P. D. A. Biology of Chromatin. In *Introduction to Epigenetics [Internet]*, DOI: 10.1007/978-3-030-68670-3_1 (Springer, 2021).
2. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492, DOI: 10.1038/nprot.2017.124 (2017).
3. Tan, J. *et al.* Cell-type-specific prediction of 3D chromatin organization enables high-throughput in silico genetic screening. *Nat. Biotechnol.* **41**, 1140–1150, DOI: 10.1038/s41587-022-01612-8 (2023).