Protein Clustering as a Mechanism in Gene Regulation:

A Simulation Study

Elena Espinosa Miñano, Supervisor: Dr. Chris Brackley

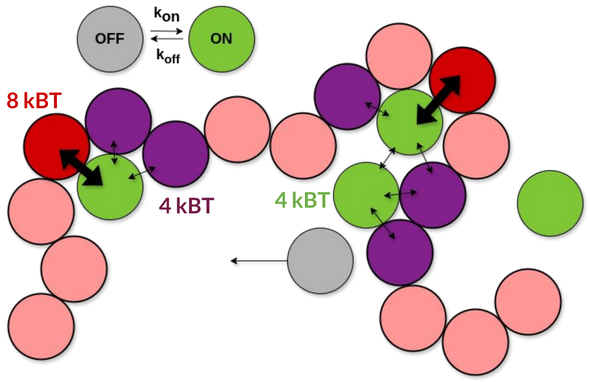
github.com/ElenaEspinosaMinano

Introduction

- The highly structured and non-random organisation of the genome is thought to be linked to cell function. DNA coils around proteins to form the **chromatin fibre** which **interacts with itself and proteins** to compact and carry out biological processes [1,2].
- **Polymer physics models** are used to simulate molecular interactions between proteins and chromatin and have helped to shed light into how chromatin-binding proteins affect and drive genome structure, protein phase separation and regulate gene expression [2].

Methods

- Chromatin fibre (5 Mbp long) and diffusing proteins are represented as different types of beads (see Fig. 1). A single bead corresponds to ~1 kbp of chromatin.
- Chromatin binding sites are determined from data in the ENCODE project [3].
- Proteins probabilistically switch "on"/"off" at a rate k_on = k_off = 2x10^(-5) inverse time-steps. When "on" ("off") they can (cannot) bind to binding sites.



	Color	Bead type	Number of beads
Chromatin (5000)		Type 1	4669
		Type 2	81
		Type 3	250
Proteins (600)	ON	Type 4	300
	OFF	Type 5	300

Figure 1: Schematic diagram of polymer physics model used to simulate the chromatin-proteins system. Polymer chain of 3 different bead types (type 1, 2 and 3). Diffusing active (type 4) and inactive (type 5) proteins can switch and interact with each other and the polymer chain. Bold, double-headed arrows represent **strong attraction (8 kBT)** active proteins (green) have to highly accessible DNA sites (red). Thin, double-headed arrows represent **weak attraction (4 kBT)** active proteins have to enhancer-promoter regions (purple) and to other active proteins (**model dependant**). Single arrow emanating from inactive protein (grey) represents diffusion from protein cluster once active protein switches to inactive. Table summarises key simulation data.

Previous work

Previous simulation studies using similar models have shown that:

1. With **protein-chromatin (p-c) attraction**; a protein binding to two sites along the chromatin (forming a loop), **induces** other proteins to go to that region (due to increase in local chromatin density) and bind there, forming a **protein cluster**. This mechanism is known as the **bridging-induced attraction**, **BIA** [4].
2. Chemical modifications (known as **PTMs**) that change the properties of proteins affecting their affinity for binding sites [5] (represented in simulations as proteins switching "on" ↔ "off"); result in **dynamic protein clusters with a self-limiting size** [6].
3. **Protein-protein (p-p) attraction** leads to the **liquid-liquid phase separation (LLPS)** of protein clusters in the system [7].

Overview

- Investigate a polymer physics model which combines **switching proteins** binding to specific sites along the chromatin region of the mouse *Sox2* gene with **varying protein-chromatin, protein-protein attraction strengths**.
- Run Molecular Dynamics simulations of several models.
- Contrast resulting measurements to provide insight into how these factors affect the **formation and dynamics of protein clusters** in chromatin.

Results + Discussion

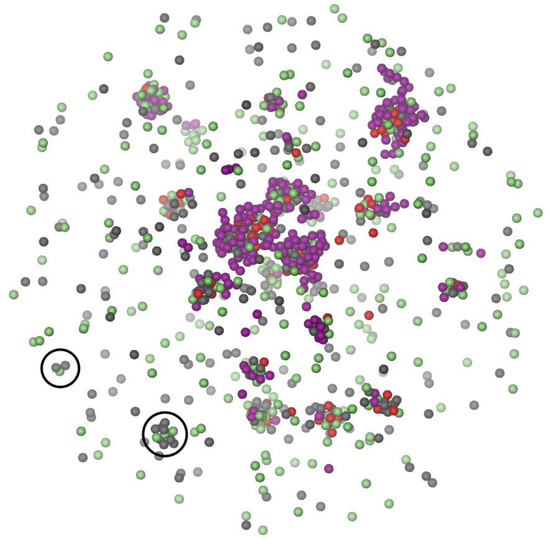
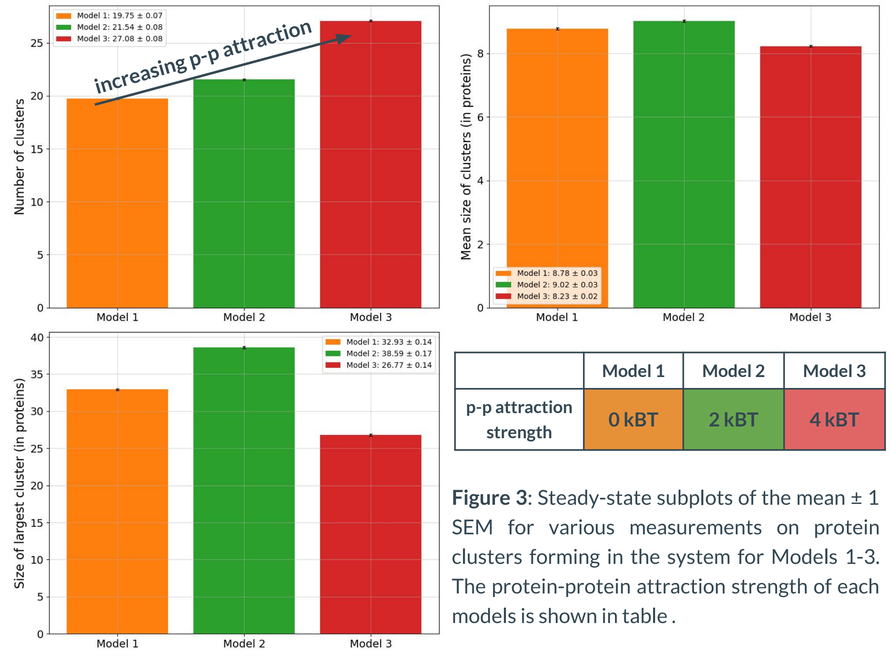


Figure 2: Snapshot of chromatin-proteins system for **Model 3**. Type 1 beads (pink in Figure 1) are not shown to facilitate seeing protein clusters. Snapshot shows **"protein-only" clusters** forming when **protein-protein attraction is 4 kBT** (circled in black).

Model 1 → 3: number of clusters increases and their mean and largest size decreases (see Fig. 3).

- In Model 3, diffusing proteins are attracted to a lot more beads (~300 active proteins) with the same attraction strength as to "weakly sticky" purple sites.
- Due to the confined system, proteins diffusing past each other, might now bind forming a real "protein-only" cluster (Fig. 2). Due to random switching, "protein-only" clusters should be small, very dynamic and short-lived. This agrees with the decrease in mean cluster size.
- In Model 1, large stretches of sticky sites are needed in order to form large protein clusters. In Model 3, only a few sticky sites are needed; as once proteins bind to these, then they themselves attract further proteins. So,



	Model 1	Model 2	Model 3
p-p attraction strength	0 kBT	2 kBT	4 kBT

Figure 3: Steady-state subplots of the mean ± 1 SEM for various measurements on protein clusters forming in the system for Models 1-3. The protein-protein attraction strength of each models is shown in table .

proteins are shared more evenly among binding sites and hence smaller clusters form.

Model 3 → 2: number of clusters decreases; mean and largest cluster size increases (Fig. 3).

- With a lower protein-protein attraction, proteins cannot form clusters on their own and need to bind to chromatin in order to cluster. It is not yet clear why and how this leads to larger clusters forming in Model 2. Further simulations are thus required.

Conclusions + Future work

- To conclude, with **switching proteins** and a **4 kBT protein-protein attraction**; **dynamic and short-lived "protein-only" clusters formed** in the chromatin region of the mouse *Sox2* gene. With a lower p-p attraction of 2 kBT these clusters did not appear.
- Repeat runs of the simulations should be conducted to validate these results.
- Further simulations are required to investigate the implications this protein clustering mechanism might have for the expression of the *Sox2* gene.

Acknowledgements

I wish to thank Chris Brackley for his patient guidance, enthusiasm and support throughout the project and in answering all my questions. Thank you to SOPA Career Development Summer Scholarships for funding this project.

References

[1] B. Alberts, A. Johnson, J. Lewis, et al. "Molecular Biology of the Cell". 4th edition. New York: Garland Science (2002).
[2] T. Misteli. "The Self-Organizing Genome: Principles of Genome Architecture and Function". *Cell* **183**, 28-45 (2020)
[3] The ENCODE Project Consortium "An integrated encyclopedia of DNA elements in the human genome". *Nature* **489**, 57–74 (2012).
[4] C. A. Brackley, S. Taylor, D. Marenduzzo, et al. "Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization". *PNAS* **110(38)**, E3605–E3611 (2013)
[5] S. Ramazi and J. Zahiri. "Post-translational modifications in proteins: resources, tools and prediction methods". *Database* **2021**, 1-20 (2021).
[6] C. A. Brackley, B. Liebchen, D. Michieletto, *et al.* "Ephemeral Protein Binding to DNA Shapes Stable Nuclear Bodies and Chromatin Domains". *Biophysical Journal* **112**, 1085-1093 (2017).
[7] M. Ancona and C. A. Brackley. "Simulating the chromatin-mediated phase separation of model proteins with multiple domains". *Biophysical Journal* **121**, 2600-2612 (2022)