Modeling chromatin fibre folding for human embryonic stem cells

FRANCESCA MAGGIONI[1], MARIDA BERTOCCHI[1], ETTORE MOSCA[2], ROLLAND REINBOLD[2] AND ILEANA ZUCCHI[2]

[1]Department of Mathematics, Statistics, Computer Science and Applications, University of Bergamo, ITALY francesca.maggioni@unibg.it

[2]2 Institute of Biomedical Technologies, ITB-CNR, Segrate Milan, ITALY

ABSTRACT

In this poster we analyze the chromatin state of Pluripotent Stem Cells by means of geometric modelling of fibre conformation. The model takes into account of the local structure of chromatin organised into *euchromatin*, permissive for gene activation, and *heterochromatin*, transcriptionally silenced. Euchromatin is assumed to be modelled by a linear DNA while heterochromatin by means of a solenoid structure in which DNA winds onto six nucleosome spools per turn.Two geometric models are presented, compared in terms of geometric quantities and tested on ChIP Data (Chromatin ImmunoPrecipitation) generated from human pluripotent embryonic stem cells. This study provides information on relationships between geometry and the transcriptional regulation in stem cells contributing to pluripotency and self-renewal.

1 Introduction

- The human genome is estimated to contain 30,000 to 40,000 unique genes. Though every gene exists within every cell in the human body, only a small percentage of genes is active in any given cell.
  What promotes the transcription of cell-specific genes and determines the cell identity? The **chromatin structure** and its ability for remodeling into different states. The chromatin state recently emerged as one of the governing factors for pluripotency of embryonic stem (ES) cells.
- Recent data show that stem cell chromatin is distinct from that of somatic or differentiated cells in several different structural and functional aspects such as **global chromatin arrangement**, condensation and compaction.
- The **chromatin state** of a cell is defined through the establishment and the maintenance of localized **open and closed states** of the **chromatin structure**, determined by **epigenomic interactors**.
- In fact **mutations** in epigenomic regulators have the potential to alter the chromatin structure, leading to mis-regulation of gene expression and contributes to cancer or other diseases.
- Therefore **understanding chromatin remodeling** is of fundamental significance in understanding cancer and for regenerative medicine.

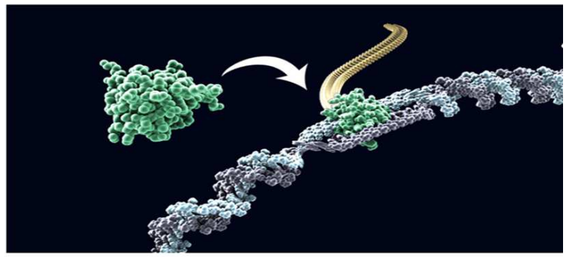


Figure: A regulatory protein that binds to DNA and affects the transcription of specific genes

2 ChIP Data (Chromatin ImmunoPrecipitation)

- To study the **chromatin state of PSCs** we utilize extensive ChIP data generated from native human pluripotent embryonic stem cells.
- **ChIP (Chromatin ImmunoPrecipitation)** is the experimental strategy used to identify the chromatin state of pluripotent stem cells (PSCs).
- ChIP is the selective enrichment of a chromatin fraction by using an antibody against a specific chromatin protein. Antibody binding can be used to determine whether local regions of chromatin are in a **heterochromatin** (highly condensed) or **euchromatin** (less extended) state in vivo.
- ChIP experiments were performed with antibodies specific for **Oct4**, **Sox2** and **Nanog**, transcription factors shown to be important for maintaining the pluripotential state of embryonic stem cells [1].

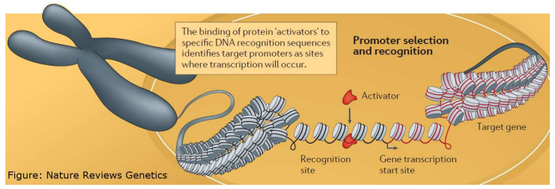


Figure: Nature Reviews Genetics

3 Geometric Models of the 30nm chromatin fiber

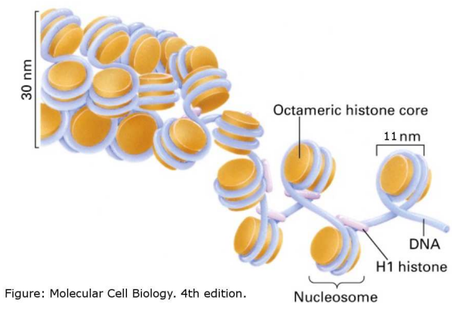


Figure: Molecular Cell Biology. 4th edition.

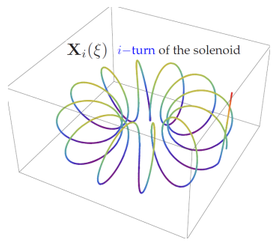
- **Assumptions:** the local structure of the chromatin is periodical [5]:
  - 6 nucleosomes per turn;
  - radius of each nucleosome: $r = 5.5$ nm;
  - wrapping angle around a nucleosome: $\theta = 3.5\pi$;
  - pitch of the helix nucleosome: $P = 2.8$ nm;
  - nucleosome wrapping length: $\Lambda = \theta\sqrt{r^2 + \frac{P^2}{4\pi^2}} \approx 5.5$ nm ≈ 147 bp;
  - linker DNA length: $b = 50$ bp (**repeat length:** $L = \Lambda + b = 197$ bp);

3.1 Torus unknot model for a solenoid turn (Heterochromatin)

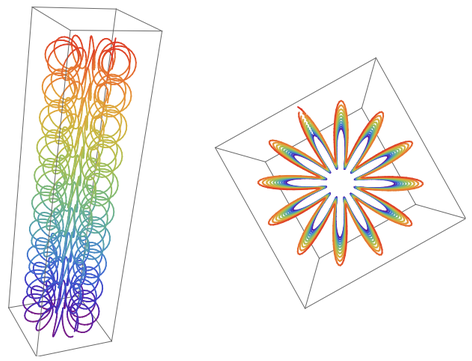
Let $\mathbf{X}_i(\xi)$ be an open **Fourier torus unknot** where $i(2\pi) \le \xi \le [(i+1)2\pi - h]$, $(i = 0, \dots, N-1)$ is a parameter along the curve.
$\mathbf{X}_i(\xi)$ represents the $i-$turn of the solenoid with N layers:

$$\mathbf{X}_i(\xi) \; : \; \begin{cases} x = Cos[\xi](R + rCos[(Q/P)\xi]) \cdot (6L - b)/l(\xi) \\ y = Sin[\xi](R + rCos[(Q/P)\xi]) \cdot (6L - b)/l(\xi) \\ z = rSin[(Q/P)\xi] \cdot (6L - b)/l(\xi) \end{cases}$$

- $R = 9.5$ nm, $r = 5.5$ nm ($R + r = 15$ nm), $P = 1$, $Q = 12$, $6L - b = 1132$bp;
- The $i-$turn $\mathbf{X}_i(\xi)$ of the solenoid is normalized by the *length function* $l_i(\xi) = l(\xi)$ s.t. its length is fixed at $6L - b = 1132$ bp.
- The $i-$turn $\mathbf{X}_i(\xi)$ of the solenoid is connected with the $(i+1)-$turn $\mathbf{X}_{i+1}(\xi)$ by means of **cubic hermite spline functions** of length b (see **Section 4**).



Example of Solenoid Model (10 turns)

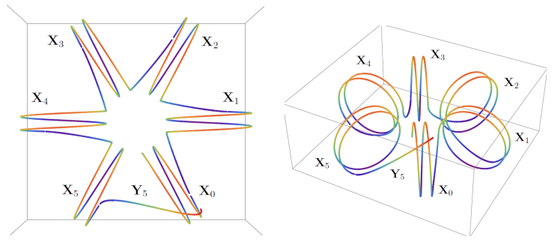


3.2 Six helices model for a solenoid turn (Heterochromatin)

Nucleosome i: $\quad i = 0, \dots, 5 \qquad 2i(2\pi) \le \xi \le (2i+1)2\pi$

$$\mathbf{X}_i \begin{cases} x = \frac{\pi\left(P - 2r\sin\left(\frac{1}{6}(2\pi i + \pi)\right)\cos(\theta(\xi - 4\pi i))\right) + P(-4\pi i + \xi - \pi)\cos\left(\frac{1}{6}(2\pi i + \pi)\right) + 2\pi R\cos\left(\frac{1}{3}\pi(i-1)\right)}{2\pi} \cdot \frac{\Lambda}{l(\xi)} \\ y = \frac{P(-4\pi i + \xi - \pi)\sin\left(\frac{1}{6}(2\pi i + \pi)\right) + 2\pi r\cos\left(\frac{1}{6}(2\pi i + \pi)\right)\cos(\theta(\xi - 4\pi i)) + 2\pi R\sin\left(\frac{1}{3}\pi(i-1)\right)}{2\pi} \cdot \frac{\Lambda}{l(\xi)} \\ z = r\sin(\theta(\xi - 4\pi i)) \cdot \frac{\Lambda}{l(\xi)} \end{cases}$$

- The $i-$nucleosome $\mathbf{X}_i(\xi)$ of a solenoid turn is normalized by the *length function* $l(\xi)$ s.t. its length is fixed at $\Lambda = 147$ bp.



4 Cubic spline functions for linkers DNA

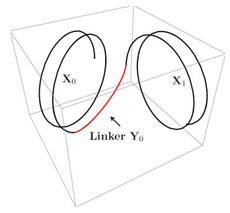
- Connections $\mathbf{Y}_i(\xi)$, with $2\pi(2i+1) \le \xi \le 2\pi(2i+2)$, $(i = 0, \dots, 5)$ between two succesive nucleosomes are approximated by **cubic Hermite spline interpolation**, a third-degree spline with each polynomial as follows:

$$\mathbf{Y}_i \begin{cases} x = a_{00}^{(i)} h_{00}(\xi) + a_{01}^{(i)} 2\pi h_{01}(\xi) + a_{10}^{(i)} 2\pi h_{10}(\xi) + a_{11}^{(i)} h_{11}(\xi)\,, & i = 0, \dots, 5 \\ y = b_{00}^{(i)} h_{00}(\xi) + b_{01}^{(i)} 2\pi h_{01}(\xi) + b_{10}^{(i)} 2\pi h_{10}(\xi) + b_{11}^{(i)} h_{11}(\xi)\,, & i = 0, \dots, 5 \\ z = c_{00}^{(i)} h_{00}(\xi) + c_{01}^{(i)} 2\pi h_{01}(\xi) + c_{10}^{(i)} 2\pi h_{10}(\xi) + c_{11}^{(i)} h_{11}(\xi)\,, & i = 0, \dots, 5 \end{cases}$$

where

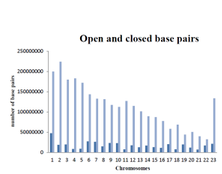
$$\begin{cases} h_{00}(\xi) &= 2\xi^3 - 3\xi^2 + 1 \\ h_{01}(\xi) &= -2\xi^3 + 3\xi^2 \\ h_{10}(\xi) &= \xi^3 - 2\xi^2 + \xi \\ h_{11}(\xi) &= \xi^3 - \xi^2 \end{cases}$$

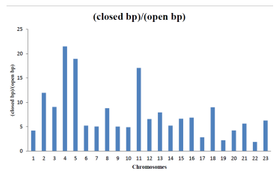
are Hermite basis functions and $a_{00}^{(i)}$, $b_{00}^{(i)}$, $c_{00}^{(i)}$, $a_{11}^{(i)}$, $b_{11}^{(i)}$, $c_{00}^{(i)}$, respectively the stating and final point with their derivatives $a_{10}^{(i)}$, $b_{10}^{(i)}$, $c_{10}^{(i)}$ and $a_{01}^{(i)}$, $b_{01}^{(i)}$, $c_{01}^{(i)}$ in the $i-$linker $(i = 0, \dots, 5)$.



5 Numerical results (ChIP Data Analysis)

5.1 Open and Closed base pairs





- In PSCs the chromatin state is **not uniform**, with **Oct4**, **Sox2** and **Nanog** promoting the most decondensation of chromatin structure in **ch. 22**.

5.2 Comparative analysis: measures and energetics of filament coiling

Let $\mathscr{C}$ be a smooth, simple curve in $\mathbb{R}^3$ given by $\mathbf{X}(\xi) : [0, L_{fin}] \longrightarrow \mathbb{R}^3$, with curvature $c(\xi)$ and torsion $\tau(\xi)$ where ξ is a parameter along the curve, $\dot{\mathbf{t}}(\xi) \equiv \mathbf{X}'(\xi)/\|\mathbf{X}'(\xi)\|$ is the unit tangent to $\mathscr{C}$ at ξ and $l(\xi)$ the length function.

We consider the following quantities:

- **normalized total curvature**

$$\mathcal{K} := \frac{1}{l(\xi)} \int_{\mathscr{C}} c(\xi)\, \|\mathbf{X}'(\xi)\|\, \mathrm{d}\xi$$

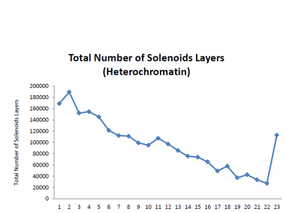
- **normalized bending energy**

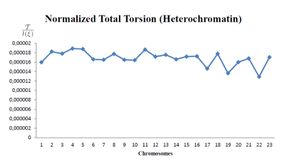
$$E_b := \frac{1}{l(\xi)} \int_{\mathscr{C}} (c(\xi))^2 \|\mathbf{X}'(\xi)\|\, \mathrm{d}\xi$$

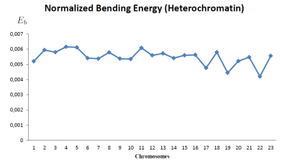
- **normalized total twist number**

$$Tw := \frac{1}{l(\xi)} \left[\frac{1}{2\pi} \int_{\mathscr{C}} \tau(\xi) \|\mathbf{X}'(\xi)\|\, \mathrm{d}\xi + \frac{1}{2\pi} [\Theta]_{\mathscr{F}} \right] = \frac{1}{l(\xi)} (\mathcal{T} + \mathcal{N})$$

where $\frac{\mathcal{T}}{l(\xi)}$ is the normalized total torsion and $\frac{\mathcal{N}}{l(\xi)}$ the normalized intrinsic twist.







	$\mathcal{K}$ per turn	$\mathcal{T}$ per turn	E_b per turn	base pairs per turn ($= 6L$)
torus unknot model	75.4342	-12.4605	5.0076	1182
6 helices model	76.9785	0.1467	7.6170	1182

6 Conclusions and References

- Two **geometric models for chromatin fibre conformation** are presented, compared in terms of geometric quantities and tested on **ChIP Data** generated from **human pluripotent embryonic stem cells**.
- Our model provides insight into the **chromatin structure** state of human pluripotent stem cells (PSCs) in terms of **geometric information**.
- We have found that in PSCs the **chromatin state is not uniform** for all the chromosomes, with **Oct4**, **Sox2** and **Nanog** activity promoting the most decondensation of chromatin structure in **chromosomes 19** and **22** (open state).
- It was demonstrated that pluripotency can be induced in non-pluripotent cells with defined factors, that include **Oct4** and **Sox2** [4] by a process involving chromatin structural remodeling that results in cell reprogramming.
- Modeling, characterizing and defining chromatin structure is important for identifying **epigenomic agents** that may provide therapies for cancer [3] and other diseases and great promise in the field of regenerative medicine [2].
- Ongoing work with ChIP analysis, using antibodies against specific components of the **transcriptional machinery** will allow to define at **higher resolution** the structural chromatin state of specific cell types.

References

[1] Boyer et al. (2005) *Cell*, **122**.

[2] Kwa et al. (2011) *Drug Discov Today.*, **16**.