Ordered Binding in Gene Regulation

Nicholas Hilgert[1,2], John Biddle[2], and Jeremy Gunawardena[2]

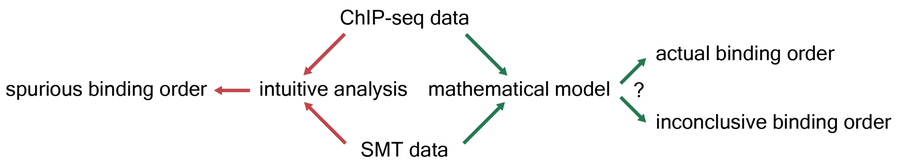
[1] Purdue University, Department of Physics and Astronomy, [2] Harvard Medical School, Department of Systems Biology

Background

During gene regulation, do transcription factors (TFs) have a temporal binding order? This is a question in the ongoing investigation of so-called “pioneer factors” that prime chromatin for further binding. Such factors might play a critical role in inducing pluripotency.

Experimental techniques such as ChIP-seq and single-molecule tracking (SMT) have previously been used to infer binding order, though proper analysis of their data requires more sophisticated methods than previously used.

We interpret data generated by these techniques on the premise of a new mathematical model founded on fundamental, physics-based assumptions.

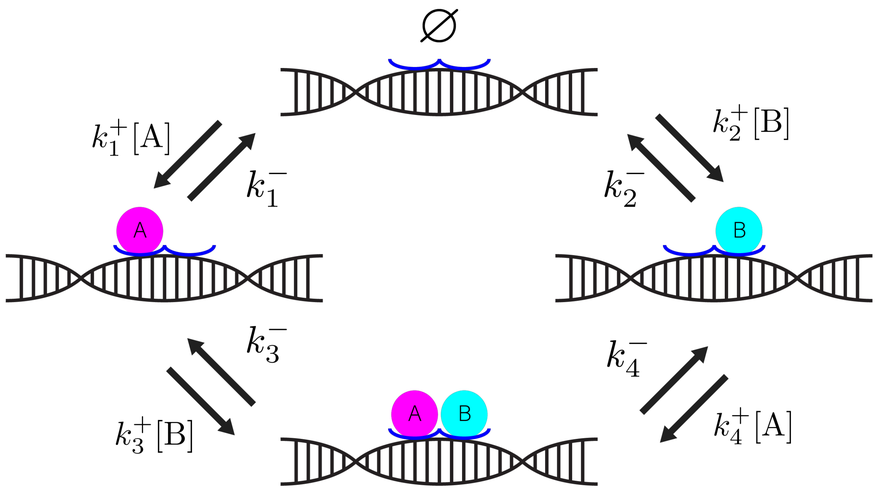


The model will determine if purported binding order conclusions from experiments are valid.

Mathematical Framework

We consider pairs of TFs through a graph-theoretic representation of a Markov process on four binding states.

The graph, for TFs A and B:

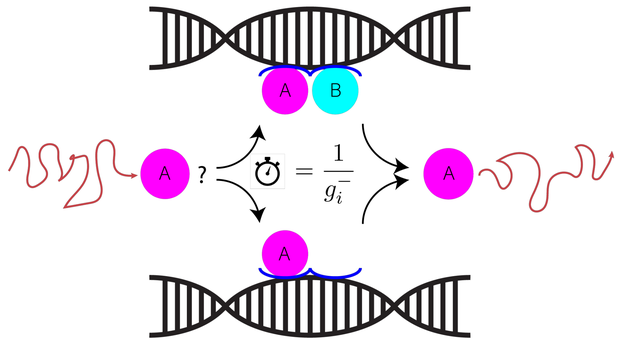


A quantitative measure of order is defined as a probabilistic preference in traversing one path over the other.

$$\frac{\mathbb{P}(\varnothing \to \mathrm{A} \to \mathrm{AB})}{\mathbb{P}(\varnothing \to \mathrm{B} \to \mathrm{AB})} = \frac{k_1^+ k_3^+ \left(k_2^- + k_4^+[\mathrm{A}]\right)}{k_2^+ k_4^+ \left(k_1^- + k_2^+[\mathrm{B}]\right)}$$

Analysis: Single-Molecule Tracking

Chen et al. (2014) use SMT to directly observe TF association and dissociation rates $g_i^{\pm}$ as the reciprocals of residence times, but only one TF is watched at once. It is unclear if the TF binds to a site already occupied by a cooperative factor.



SMT data gives a set of $g_i^{\pm}$ for the TFs Sox2 and Oct4. We seek a mapping between measured rates $g_i^{\pm}$ corresponding to the TF and graph rates $k_i^{\pm}$ corresponding to the DNA kinetics that determine binding order. Chen et al. (2014) erroneously assume $g_i^{\pm} = k_i^{\pm}$.

$$\left\{ \begin{array}{c} g_1^+ \\ g_1^- \\ \vdots \\ g_4^- \end{array} \right\} \Longleftrightarrow \left\{ \begin{array}{c} k_1^+ \\ k_1^- \\ \vdots \\ k_4^- \end{array} \right\}$$

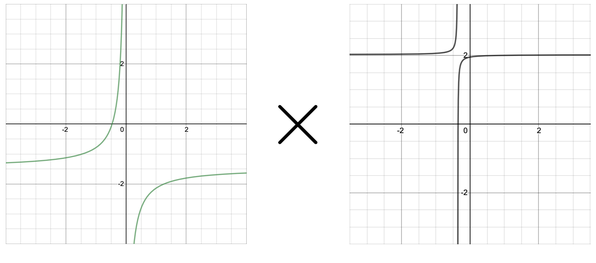
We treat the experimental residence time as a Markov first hitting time, from which we derive an algebraic relationship between the $g_i^{\pm}$ and $k_i^{\pm}$ to match the experimental protocol from Chen et al. (2014).

$$\frac{1}{g_4^+[\mathrm{DNA}]} = \frac{k_1^- \mathbb{P}(\mathbf{s})}{k_1^- \mathbb{P}(\mathbf{s}) + k_4^- \mathbb{P}(\mathbf{so})} \cdot \frac{k_2^- + k_2^+[\mathrm{O}_i] + k_4^+[\mathrm{S}_f]}{k_2^- k_1^+[\mathrm{S}_f] + k_1^+[\mathrm{S}_f] k_4^+[\mathrm{S}_f] + k_4^+[\mathrm{S}_f] k_2^+[\mathrm{S}_f]} + \frac{k_4^- \mathbb{P}(\mathbf{so})}{k_1^- \mathbb{P}(\mathbf{s}) + k_4^- \mathbb{P}(\mathbf{so})} \cdot \frac{k_2^- + k_2^+[\mathrm{O}_i] + k_1^+[\mathrm{O}_i]}{k_2^- k_1^+[\mathrm{S}_f] + k_1^+[\mathrm{S}_f] k_4^+[\mathrm{S}_f] + k_4^+[\mathrm{S}_f] k_2^+[\mathrm{S}_f]}$$

Repeating the process for all rates generates a system of polynomial equations whose varieties are analyzed through techniques from algebraic geometry. Preliminary analysis shows that these relationships are sensitive and multi-valued.

$$b(a\delta(\alpha+\beta) + \beta c(\mathbf{x}-\alpha) - \alpha\delta(\beta+\mathbf{x})) + \beta(\delta+\mathbf{z})(-\alpha(a+c+d) + ad + a\mathbf{x} + c\mathbf{x}) \times ((a+c)\mathbf{xz} + bc\mathbf{x} + ad\mathbf{z})$$

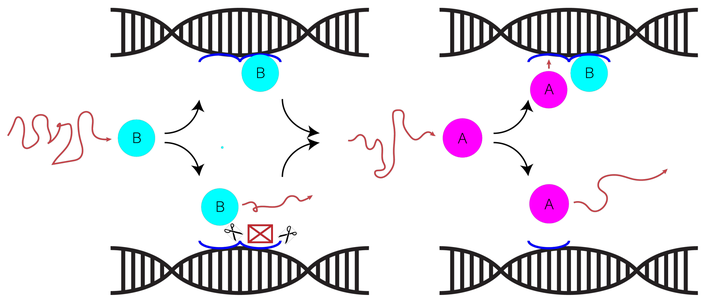
$=$



Analysis: ChIP-seq

ChIP-seq data gives a set of probability distributions of binding states, as used in Xie et al. (2017) to infer binding interdependence between TFs Sox2, Oct4, Klf4, and Esrrb.

In Xie et al. (2017), TF binding sites are deleted to prohibit potentially cooperative binding of TF partners.

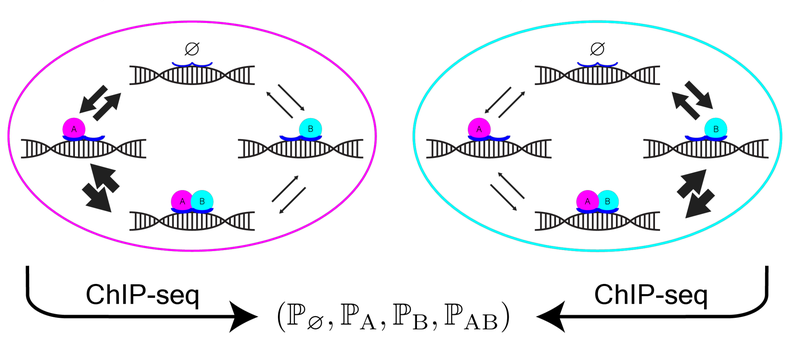


If ChIP-seq is a successful experimental technique, a measured probability distribution should imply a unique binding order.

$$(\mathbb{P}_\varnothing, \mathbb{P}_\mathrm{A}, \mathbb{P}_\mathrm{B}, \mathbb{P}_\mathrm{AB}) \Longleftrightarrow \frac{\mathbb{P}(\varnothing \to \mathrm{A} \to \mathrm{AB})}{\mathbb{P}(\varnothing \to \mathrm{B} \to \mathrm{AB})}$$

With the Matrix-Tree Theorem, we show that a family of graphs with different orders can be mapped to identical probability distributions measured by ChIP-seq.

$$\frac{\mathbb{P}(\varnothing \to \mathrm{A} \to \mathrm{AB})}{\mathbb{P}(\varnothing \to \mathrm{B} \to \mathrm{AB})} > 1 \qquad \frac{\mathbb{P}(\varnothing \to \mathrm{A} \to \mathrm{AB})}{\mathbb{P}(\varnothing \to \mathrm{B} \to \mathrm{AB})} < 1$$



It follows that ChIP-seq on its own is not sufficient to conclude a binding order in this context.

Discussion

Concentration dependence of the graph edge labels and corresponding kinetics suggests a highly context-dependent relationship between binding order and TFs, likely not an inherent property of the factors themselves.

A groundbreaking in vitro study by Li et al. (2019) uses both fluorescence and SMT to directly observe a binding order for Sox2 and Oct4, but it is unclear if its conclusions generalize to in vivo settings.

The task of inferring binding order is more difficult than it initially seems, and current experiments are potentially deficient for doing so.

Acknowledgements

Thanks to John Biddle, Jeremy Gunawardena, and the Gunawardena lab for their excellent mentorship throughout the project, as well as my intern cohort for their friendship and camaraderie.

References

1. J. Biddle et al. Negative reciprocity, not ordered assembly, underlies the interaction of Sox2 and Oct4 on DNA. *eLife*, 2019.
2. J. Chen et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*, 2014.
3. S. Li, E. B. Zheng, L. Zhao, and S. Liu. Nonreciprocal and conditional cooperativity directs the pioneer activity of pluripotency transcription factors. *bioRxiv*, 2019.
4. L. Xie et al. A dynamic interplay of enhancer elements regulates klf4 expression in nave pluripotency. *Genes & development*, 2017.