The Role of Cooperative Binding in Regulation of Gene Expression

Samuel A. Ocko[1]

[1]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*

Gene expression can be regulated through transcription factors, which bind to specific sites on DNA and allow for transcription into RNA. When transcription requires several transcription factors to each be bound to a site, interactions between transcription factors can lead to different kinds of dependences of transcription rates on transcription factor concentration. In this paper we examine the role of interaction between transcription factors in gene expression. We introduce a general model of interacting neighboring transcription factors which can be solved using transfer matrix techniques. We then examine a specific model which we show to be equivalent to the 1D Ising model, and show that cooperative binding leads to a sharper dependence of gene expression on transcription factor concentration.

Regulation of gene expression is essential for the ability of a cell to respond to changes in it's environment. One way that gene expression is regulated is through concentrations of transcription factors, which affect the rate at which a gene is transcribed. For certain segments, transcription factors must be bound to specific binding sites on DNA before transcription of a genetic sequence can occur. Higher concentrations of a transcription factor lead to a higher probability of a site being bound to that transcription factor. Thus, higher concentrations of a transcription factor will lead to higher transcription rates and gene expression.

Making gene expression be regulated through the binding of a single transcription factor to a single binding site has serious limitations; there is only one possible way for the rate of transcription to depend on the concentration of the transcription factor. This problem can be somewhat mitigated through having several binding sites which bind independently, allowing transcription only when *all* binding sites are bound by appropriate transcription factors. However, the transcription rate is then still limited to being the product of single-transcription factor terms.

More complicated behavior can be achieved through having interactions *between* transcription factors, known as cooperative binding. With interactions, the binding or non-binding of a certain site can raise or lower the probability of binding at neighboring sites. In this paper we present a simple model to describe these phenomena. We first show our model to be equivalent to one-dimensional spin systems in statistical physics, and will show how to solve the the model using transfer matrix techniques. As an example, we will give a set of couplings which are equivalent to the 1D Ising model and solve it, showing how interactions between transcription factors affect the rate of gene expression.

Simple Models: In the simplest possible model, the expression of a gene is regulated by a single binding site which is either bound or unbound to a transcription factor. When the transcription factor has a concentration of ρ in the surrounding fluid, the net change in free energy when a transcription factor is bound is:

$$\mathbf{E} = -\mathcal{A} - k_B T \ln \rho \tag{1}$$

Where $\mathcal{A}$ is the affinity the transcription factor has for the binding site. We also have a concentration-dependent term, which represents the change of free energy in the surrounding fluid due to it having one less transcription factor. The probability that a transcription factor is bound is thus:

$$\frac{e^{-\beta \mathbf{E}}}{e^{-\beta \mathbf{E}} + 1} = \frac{\rho/\rho^c}{\rho/\rho^c + 1},$$

where we have defined the characteristic concentration ρ^c to be $e^{-\beta \mathcal{A}}$.

We may extend the model to include $\mathcal{N}$ binding sites, each of which binds to a possibly different transcription factor. Transcription occurs only when *every* site is bound by a transcription factor. When there are no interactions between adjacent binding sites, the transcription rate is simply proportional to the product of individual binding probabilities:

$$\prod_{i=1}^{\mathcal{N}} \frac{\rho_i/\rho_i^c}{\rho_i/\rho_i^c + 1}. \tag{2}$$

However, there are often interactions between adjacent binding sites on a complex, where the free energy is lower if adjacent sites are both bound or both unbound. If we assume this interaction to be very strong, the complex effectively has only two possible states; one where every site is bound and one where no sites are bound. This will give us the activation probability:

$$\frac{\prod_{i=1}^{\mathcal{N}} \rho_i/\rho_i^c}{\prod_{i=1}^{\mathcal{N}} \rho_i/\rho_i^c + 1}. \tag{3}$$

This is known as the Hill equation[1], originally developed to model the cooperative binding in hemoglobin.

While Eq.(2), Eq.(3) give us the ability to calculate transcription rates for very strong interactions and very weak interactions, we would like to be able to calculate transcription rates for a more general model of binding.

General Model of Binding: The expression of a gene is regulated by $\mathcal{N}$ binding sites, each of which is empty or has a transcription factor bound to it. The state of the system is described by the $\mathcal{N}$ -dimensional bit vector $\vec{\alpha}$, where $\alpha_i = 0, 1$ defines the occupation of site i. The system is in a statistical ensemble of states $\vec{\alpha}$, and transcription is possible only when *all* sites are bound by an appropriate transcription factor, $\vec{\alpha}_{exp} = (1, 1, \dots 1)$. The rate of gene expression is proportional to the probability that all sites are bound by transcription factors: $R_{\mathcal{G}} \propto P(\vec{\alpha}_{exp})$. We would like to find $P(\vec{\alpha}_{exp})$.

The probability of the system being in state $\vec{\alpha}$ is determined by the free energy of $\vec{\alpha}$: $P(\vec{\alpha}) \propto e^{-\beta E_f(\vec{\alpha})}$. The free energy of state $\vec{\alpha}$ is given by the single-body terms $\mathbf{E}_i$ and two-body terms $\mathbf{E}^i$:

$$E_f(\vec{\alpha}) = \sum_{i=1}^{\mathcal{N}} \alpha_i \mathbf{E}_i + \mathcal{E}^i_{\alpha_i, \alpha i+1}.$$

$\mathbf{E}_i$ is the change in free energy from binding a transcription factor to site i, and $\mathcal{E}^i_{\alpha_i, \alpha i+1}$ is the interaction free energy between adjacent binding sites $i, i+1$ [4].

We note that this model is equivalent to 1D spin models; a bound site is equivalent to an up spin, and an unbound site is equivalent to a down spin. The single-body free energy $\mathbf{E}_i$ becomes the magnetic field applied to that spin, and the interaction term $\mathcal{E}^i_{\alpha_i, \alpha i+1}$ becomes an interaction term between adjacent spins. To solve this system, we can use the same techniques that are used for 1D spin models.

Transfer Matrix Technique: To solve for $P(\vec{\alpha}_{exp})$, it is insufficient to simply know $E_f(\vec{\alpha}_{exp})$; we must also know the partition function. We may write the partition function as:

$$\mathcal{Z} = \sum_{\vec{\alpha}} e^{-\beta E_f(\vec{\alpha})} = \sum_{\alpha_1 \dots \alpha_{\mathcal{N}}} \prod_i e^{-\beta(\alpha_i \mathbf{E}_i) + \mathcal{E}^i_{\alpha_i, \alpha i+1}}.$$

Defining the two-dimensional "transfer matrix"

$$\mathcal{M}^i_{\alpha_i, \alpha_{i+1}} = e^{-\beta\left(\frac{1}{2}\mathbf{E}_i(\alpha_i + \alpha_{i+1}) + \mathcal{E}^i_{\alpha_i, \alpha i+1}\right)},$$

we note that we may rewrite the partition function as

$$\mathcal{Z} = \sum_{\alpha_1 \dots \alpha_{\mathcal{N}}} \mathcal{M}^1_{\alpha_1, \alpha_2} \mathcal{M}^2_{\alpha_2, \alpha_3} \dots \mathcal{M}^{\mathcal{N}-1}_{\alpha_{\mathcal{N}-1}, \alpha_{\mathcal{N}}} \mathcal{M}^{\mathcal{N}}_{\alpha_{\mathcal{N}}, \alpha_1}.$$

Interpreting this expression as a matrix multiplication, we write the partition function as the trace of the product of the transfer matrices:

$$\mathcal{Z} = \mathrm{Tr}\left[\mathcal{M}^1 \mathcal{M}^2 \dots \mathcal{M}^{\mathcal{N}}\right].$$

Given any set of interactions, we can easily first construct the transfer matrices $\mathcal{M}^i$, and from those the partition function $\mathcal{Z}$. It is then straightforward to calculate the probability that all sites are bound:

$$P(\vec{\alpha}_{exp}) = \frac{e^{-\beta E_f(\vec{\alpha}_{exp}})}{\mathcal{Z}}$$

This equation, in principle, gives us a method for computing $P(\vec{\alpha}_{exp})$ for any kind of interaction between neighboring transcription factors. We would like to increase our understanding through looking at a simple example.

Simple Example: We will assume a simple model where all sites act equivalently, and prove it to be equivalent to the 1D Ising model. Every binding site will bind to the same transcription factor with the same affinity. For interactions, free energy will be lowered by J when neighboring sites have the same binding state and will be raised by J when neighboring sites have opposite binding states:

$$H = \sum_i \alpha_i \mathbf{E} - J(2\alpha_i - 1)(2\alpha_{i+1} - 1)$$

This is equivalent to the one-dimensional classical Ising model of $\mathcal{N}$ spins, where a bound site is equivalent to an up spin, and an unbound site is equivalent to a down spin:

$$H_{Is} = \mathbf{E} \sum_i \frac{1 + Z_i}{2} - J \sum_i Z_i \cdot Z_{i+1}.$$

This gives us a transfer matrix of

$$\mathcal{M} = \begin{pmatrix} e^{-\beta(-J+\mathbf{E})} & e^{-\beta(J+\mathbf{E}/2)} \\ e^{-\beta(J+\mathbf{E}/2)} & e^{-\beta(-J)} \end{pmatrix},$$

having eigenvalues of:

$$\lambda_{\pm} = \frac{1}{2} e^{\beta J}\left(e^{-\beta \mathbf{E}} + 1\right) \pm \sqrt{\left(e^{-\beta(J+\mathbf{E}/2)}\right)^2 + \left(\frac{1}{2} e^{\beta J}(e^{-\beta \mathbf{E}} - 1)\right)^2}.$$

While the system actually occupies a segment on a strand of DNA and is not periodic, for the sake of mathematical simplicity we will assume the system to be periodic, giving a partition function of

$$\mathcal{Z} = \mathrm{Tr}\left[\mathcal{M}^{\mathcal{N}}\right] = \lambda_+^{\mathcal{N}} + \lambda_-^{\mathcal{N}},$$

and a probability of activation of

$$P(\vec{\alpha}_{exp}) = \frac{\left(e^{-\beta \mathbf{E} + \beta J}\right)^{\mathcal{N}}}{\lambda_+^{\mathcal{N}} + \lambda_-^{\mathcal{N}}} \approx \left(\frac{e^{-\beta \mathbf{E} + \beta J}}{\lambda_+}\right)^{\mathcal{N}}.$$

We plot the the dependence on concentration for several values of J for a system of 5 sites to demonstrate the behavior.

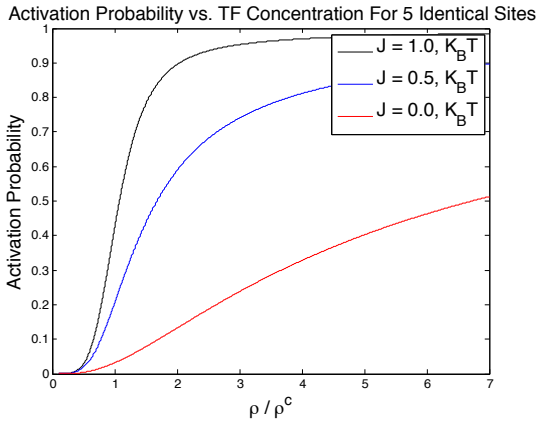

FIG. 1: Cooperative binding yields a steeper response curve than non-cooperative binding.

In the non-interacting limit and strongly interacting limits, we can take the limits

$$\mathcal{M}_{J\to 0} = \begin{pmatrix} \rho/\rho^c & \sqrt{\rho/\rho^c} \\ \sqrt{\rho/\rho^c} & 1 \end{pmatrix}, \quad \mathcal{M}_{J\to\infty} = e^{\beta J} \begin{pmatrix} \rho/\rho^c & 0 \\ 0 & 1 \end{pmatrix},$$

to recover Eq.(2), Eq.(3).

Extensions and Open Issues: While we have written down a model for interactions between transcription factors and solved it exactly, there is still more to do. Our model can be used for other binding systems neither in the strongly-interacting nor weakly-interacting limits, and can easily be extended to allow low-affinity and high-affinity states [2]. While our model is reasonable, determining what the parameters actually are is a difficult inverse problem.

A serious limitation is that we have assumed the system to be at equilibrium, which may be a very poor assumption. Because the energy landscape is more complicated when there are interacting transcription factors, we expect that with strong interactions, the binding complexes may take a long time to reach equilibrium and may even demonstrate hysteresis effects. Like ferromagnets, which have the same equilibrium properties as our model, binding complexes may be able to persist for a long time with all sites bound or all sites unbound. One possible way to model kinetics is to use a Markov chain model: we could use a model of target site location[3] to find an association rate, and then calculate dissociation rates through equilibrium considerations.

Acknowledgements: We thank David Liu for helpful discussions.

[1] A. V. Hill, J Physiol (Lond) **40** (1910).
[2] J. Monod, J. Wyman, and J.-P. Changeux, Journal of Molecular Biology **12**, 88 (1965).
[3] O. G. Berg, R. B. Winter, and P. H. Von Hippel, Biochemistry **20**, 6929 (1981).
[4] We are working modulo $\mathcal{N}$, so $\mathcal{E}^{\mathcal{N}}$ may be zero if we don't have periodic boundary conditions.