CSM2024: Homework 2 (85 points)

1. (25 points) **Ligand binding with depletion**

   (a) (10 points) Derive an expression for the fractional occupancy of a receptor R as a function of the total ligand and receptor concentrations, L_{tot} and R_{tot} respectively, and the binding dissociation constant K_d *without* assuming that ligand is in excess.

   Let $[RL]$ denote the amount of bound ligand-receptor complex, $[R]$ denote the amount of free receptor, and $[L]$ denote the amount of free ligand.
   The chemical process being modeled is as follows:

   $$R + L \rightleftharpoons RL$$

   We know that the total amount of R (R_{tot}) and the total amount of L (L_{tot}) are the sum of the amount of bound species (RL) and free species (R and L respectively):

   $$R_{tot} = [RL] + [R]$$

   $$L_{tot} = [RL] + [L]$$

   These statements imply the following facts:

   $$[R] = R_{tot} - [RL]$$

   $$[L] = L_{tot} - [RL]$$

   We know that the dissociation constant (K_d) is formulated as:

   $$K_d = \frac{[R][L]}{[RL]}$$

   Using the fact that $[R] = R_{tot} - [RL]$ and $[L] = L_{tot} - [RL]$ as derived above, we have that:

   $$K_d = \frac{(R_{tot} - [RL])(L_{tot} - [RL])}{[RL]}$$

   $$K_d[RL] = R_{tot}L_{tot} - (R_{tot} + L_{tot})[RL] + [RL]^2$$

   Subtracting $K_d[RL]$ from both sides gives us a quadratic in $[RL]$:

   $$0 = [RL]^2 - (R_{tot} + L_{tot} + K_d)[RL] + R_{tot}L_{tot}$$

   Using the quadratic equation, we can solve for $[RL]$:

   $$[RL] = \frac{(R_{tot} + L_{tot} + K_d) \pm \sqrt{(R_{tot} + L_{tot} + K_d)^2 - 4R_{tot}L_{tot}}}{2}$$

   The fractional occupancy is given by the ratio of bound receptor over total receptor ($\frac{[RL]}{R_{tot}}$), so the expression for the fractional occupancy becomes:

   $$\frac{[RL]}{R_{tot}} = \frac{(R_{tot} + L_{tot} + K_d) \pm \sqrt{(R_{tot} + L_{tot} + K_d)^2 - 4R_{tot}L_{tot}}}{2R_{tot}}$$

(b) (5 points) Use this expression to make a plot of the fractional error in the binding probability from assuming ligand is in excess as a function of L_{tot}. Take $R_{\text{tot}} = K_d = 50$ nM. Interpret your result.

***Please see .ipynb file for code used to create plot**

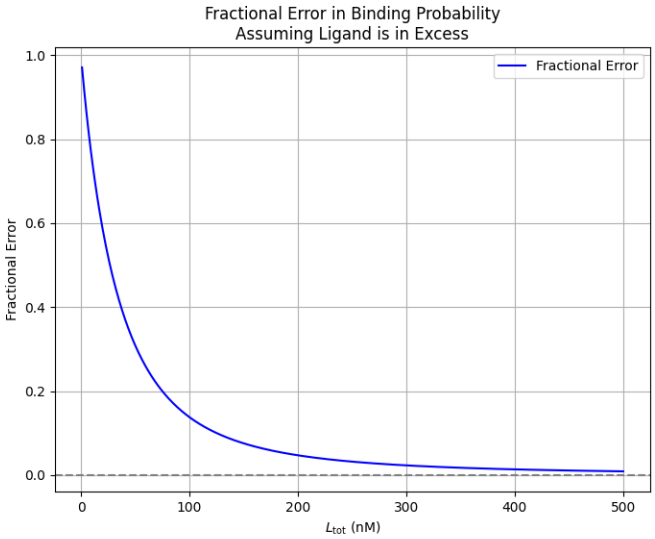

Figure 1: Plot showing the fractional error in binding probability as a function of L_{tot}.

The figure shows that at low L_{tot}, the fractional error is high because the assumption that the ligand is in excess is not valid. Thus, there is a higher error / deviation from this assumed state. However, at high L_{tot} (the situation where $L_{tot} >> R_{tot}$), the error decreases since the assumption of excess ligand becomes more accurate.

(c) (10 points) Fit the binding curves measured by Maeda et al. (2000) for binding of σ^{70} to RNA polymerase assuming the σ^{70} concentration is 0.4 nM using the expression you derived in part (a). Also fit the same data assuming that L is in excess. Plot the fits and report the values of K_d you obtain in each case. Note that the receptor in this case is σ^{70} and the ligand concentrations are given in nM. What are your conclusions?

***Please see .ipynb file for code used to create plot**

The K_d values obtained under each set of assumptions differ; under the assumption that ligand is NOT in excess, we get $K_d = 0.21$ nM, whereas under the assuption that ligand is in excess, we get $K_d = 0.15$ nM.
The excess ligand assumption is not suitable for this model , due to the relative concentrations of receptor and ligand in the data. This fact is confirmed even after fitting the data; we see that the excess ligand assumption leads to an underestimation of K_d. The true model gives us $K_d = 0.21$ nM, which is close to the K_d reported by Maeda et al. ($K_d = 0.26$ nM), whereas the excess ligand model gives $K_d = 0.15$ nM. So, we see that the excess ligand assumption leads to a more inaccurate fit and is thus unsuitable for this dataset.

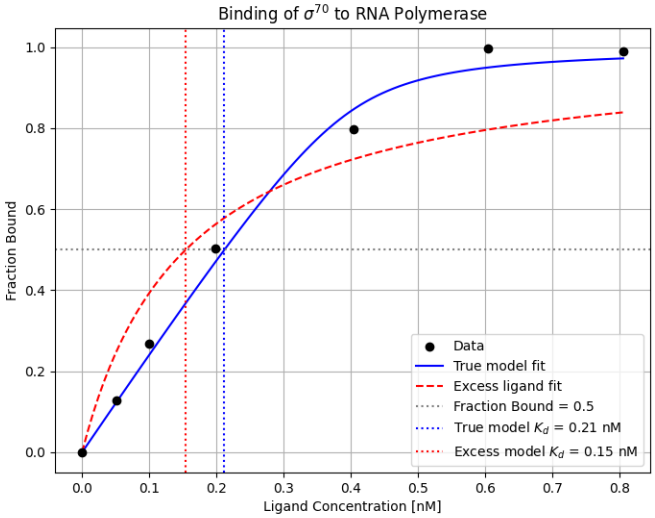

Figure 2: Plot showing the fits for binding of σ^{70} to RNA polymerase under regular vs. excess ligand assumptions

2. (30 points) **Network Motif Analysis**

   (a) (15 points) Write a program to detect all 1, 2, and 3-mode motifs in a directed graph network. The input is a list of pairs of numbers indicating the source and target for each regulatory interaction. Apply your program to the E. coli network graph given here. Be sure to avoid overcounting subgraphs that occurs when a simple subgraph appears in a more complex one. In addition to your code provide a succinct description of the logic you used to find subgraphs and to avoid overcounting.

   Please see .ipynb file for code & results
   To find subgraphs, we iterate through all of the possible 1, 2, and 3-node combinations. We exclude combinations that have no connecting edges, as these are not valid subgraphs. For each valid subgraph found, we append it to the respective list of 1, 2, or 3-node motifs based on the number of edges between the nodes.
   To avoid overcounting, we initialize a set object to store motifs that have already been found. For each found motif, we first compute the canonical form of the subgraph and check whether it has been seen before. If it has not been seen, we store it as a new motif and add it to the set of found motifs. If it has been seen, it is already in the set of found motifs, and we avoid overcounting.

   (b) (15 points) Write a program to generate ER networks with a given number of nodes and edges and use it to perform an analysis of your results in part (a) to determine which subgraphs are motifs or anti-motifs. Justify your choice of criteria for selecting these.

   Please see .ipynb file for code & results
   To determine which subgraphs are motifs or anti-motifs, we compare the frequency of their appearance in the random ER network and the *E. Coli* network. For subgraphs that occur **more frequently** in the real network than the random network, we classify these as motifs. For subgraphs that occur **less frequently** than expected at random, we classify these as anti-motifs.

3. (10 points) **Linearized Positive Autoregulation** Determine the response time for a linear model of positive autoregulation given by

$$\frac{dX}{dt} = \beta + \beta_1 X - \alpha X$$

What assumptions do you need to make about the parameters?

$$\frac{dX}{dt} = \beta + \beta_1 X - \alpha X = \beta + (\beta_1 - \alpha)X$$

Setting $\frac{dX}{dt} = 0$ to solve for the steady state solution:

$$0 = \beta + (\beta_1 - \alpha)X_{st}$$

$$-\beta = (\beta_1 - \alpha)X_{st}$$

$$X_{st} = \frac{-\beta}{\beta_1 - \alpha} = \frac{\beta}{\alpha - \beta_1}$$

Define a new variable u such that it relates the value of X to its steady state value X_{st}:

$$X = u + X_{st}$$

but $X_{st} = \frac{\beta}{\alpha - \beta_1}$, so:

$$X = u + \frac{\beta}{\alpha - \beta_1}$$

Now, we substitute this relationship back into the original differential equation:

$$\frac{du}{dt} = \beta + u(\beta_1 - \alpha) + \frac{\beta}{\alpha - \beta_1}(\beta_1 - \alpha) = \beta + u(\beta_1 - \alpha) - \beta = u(\beta_1 - \alpha)$$

So, we have:

$$\frac{du}{dt} = u(\beta_1 - \alpha)$$

The variables are separable so we can solve as follow:

$$\int \frac{du}{u} = \int (\beta_1 - \alpha)dt$$

$$\ln u = (\beta_1 - \alpha)t + C$$

Set the inital condition ($t = 0$):

$$\ln u(0) = C$$

$$u(t) = Ce^{(\beta_1 - \alpha)t}$$

Recall that $X = u + X_{st} = u + \frac{\beta}{\alpha - \beta_1}$, so:

$$X(t) = Ce^{(\beta_1 - \alpha)t} + \frac{\beta}{\alpha - \beta_1}$$

Set $t = 0$:

$$X(0) = X_0 = C + \frac{\beta}{\alpha - \beta_1} \implies C = X_0 - \frac{\beta}{\alpha - \beta_1}$$

So, the solution to the differential equation is:

$$X(t) = \left(X_0 - \frac{\beta}{\alpha - \beta_1}\right) e^{(\beta_1 - \alpha)t} + \frac{\beta}{\alpha - \beta_1}$$

To solve for the response time, we plug in the values $t = t_{1/2}$ and $X(t_{1/2}) = \frac{X_{st}}{2}$:

$$\frac{X_{st}}{2} = \left(X_0 - \frac{\beta}{\alpha - \beta_1}\right) e^{(\beta_1 - \alpha)t_{1/2}} + \frac{\beta}{\alpha - \beta_1}$$

$$\frac{\beta}{2(\alpha - \beta_1)} = \left(X_0 - \frac{\beta}{\alpha - \beta_1}\right) e^{(\beta_1 - \alpha)t_{1/2}} + \frac{\beta}{\alpha - \beta_1}$$

$$\frac{\beta}{2(\beta_1 - \alpha)} = \left(X_0 - \frac{\beta}{\alpha - \beta_1}\right) e^{(\beta_1 - \alpha)t_{1/2}} \implies \frac{\frac{\beta}{2(\beta_1 - \alpha)}}{\left(X_0 + \frac{\beta}{\beta_1 - \alpha}\right)} = e^{(\beta_1 - \alpha)t_{1/2}}$$

$$\ln \frac{\frac{\beta}{2(\beta_1 - \alpha)}}{\left(X_0 + \frac{\beta}{\beta_1 - \alpha}\right)} = (\beta_1 - \alpha)t_{1/2} \implies t_{1/2} = \frac{1}{(\beta_1 - \alpha)} \ln \frac{\frac{\beta}{2(\beta_1 - \alpha)}}{\left(X_0 + \frac{\beta}{\beta_1 - \alpha}\right)}$$

Thus, the expression for the response time is:

$$t_{1/2} = \frac{1}{(\beta_1 - \alpha)} \ln \frac{\frac{\beta}{2(\beta_1 - \alpha)}}{\left(X_0 + \frac{\beta}{\beta_1 - \alpha}\right)}$$

Assumptions about the parameters: We need $x > 0$ for $\ln(x)$ to be valid, so:

$$\beta_1 - \alpha > 0 \implies \beta_1 > \alpha$$

Thus, the assumptions implied about the parameters include that $\beta_1 > \alpha$, i.e., that the production rate $\beta_1 X$ is greater than the degradation rate, $-\alpha X$.

4. (20 points) **Shaping the Pulse**

In this problem we consider an I1-FFL as shown in the figure where gene X is initially not expressed but is turned on by simple gene regulation at time $t = 0$. Assume that genes X, Y, and Z are expressed at rates β_X, β_Y, and β_Z respectively when their promoters are in the ON state and 0 otherwise. Let the thresholds for each regulatory arrow shown be K_{XY}, K_{XZ}, and K_{YZ}. Assume that the degradation rate for all three proteins has the same value α. Finally, assume that the signals regulating X and Y are present throughout.

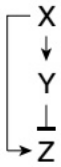

Figure 3: I1-FFL.

(a) (5 points) Make a figure showing how the pulse shape (width and height) depends on the model parameters.
Please see .ipynb file for code
In the figure, we see that the parameter β_X shifts the pulse horizontally, with increasing β_X values corresponding to rightward shifts. Moreover, we see that the parameter β_Z affects the pulse height, with higher β_Z values correlated to a higher pulse height.

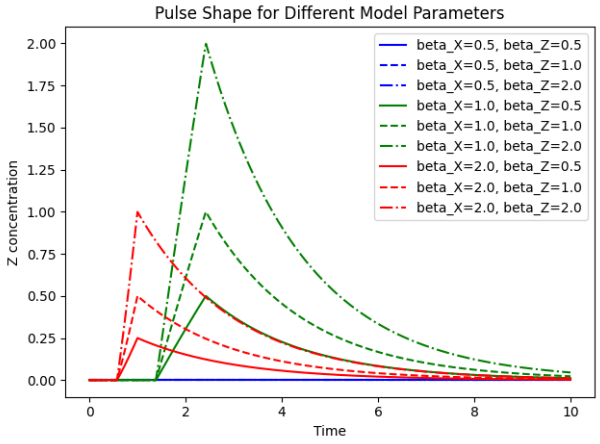

Figure 4: Pulse shapes for I1-FFL varying model parameters.

(b) (10 points) Derive expressions for the pulse width and height as a function of the model parameters. The pulse height is determined by the maximum value of Z, Z_{max}. We can assume that Z_{max} occurs at the point when the rate of production of Z balances the degradation rate of Z, i.e., when $\frac{dZ}{dt} = 0$. Recall the differential equation for Z:

$$\frac{dZ}{dt} = \beta_Z X * (X > K_{XY})(Y < K_{YZ}) - \alpha Z$$

$$0 = \beta_Z X * (X > K_{XY})(Y < K_{YZ}) - \alpha Z$$

$$\beta_Z X * (X > K_{XY})(Y < K_{YZ}) = \alpha Z$$

$$\alpha Z \propto \beta_Z X$$

$$Z \propto \frac{\beta_Z X}{\alpha}$$

The pulse width is determined by the amount of time from the beginning of production of Z, t_{start} to the end of production of Z, t_{end}. The width will depend on the degradation rate α, the production rate β_Z and β_X, and the threshold values. The amount of Z will be greatest when X is being produced maximally and Y is about to hit the threshold K_{YZ}. So, the pulse height has the following relationship:

$$Z_{max} \propto \frac{\beta_Z \cdot X}{\alpha}$$

(c) (5 points) Design thresholds such that genes Z_1 and Z_2 regulated by the same circuit turn on and off in the same order and make a figure illustrating why this design works.

Please see .ipynb file for code

The thresholds must be designed such that Z_1 turns on before Z_2, so $K_{XZ1} < K_{XZ2}$, since X activates Z. For Z_1 to turn off before Z_2, we need to design thresholds K_{YZ1} and K_{YZ2} where we know that Y inhibits Z. Z_1 must turn off first, and so the K_{YZ1} threshold must be met first, and thus $K_{YZ1} < K_{YZ2}$. Using threshold values upholding the two conditions ($K_{XZ1} < K_{XZ2}$ and $K_{YZ1} < K_{YZ2}$), we obtain the following figure:

We see that Z_1 turns on before Z_2 and Z_1 turns off before Z_2.

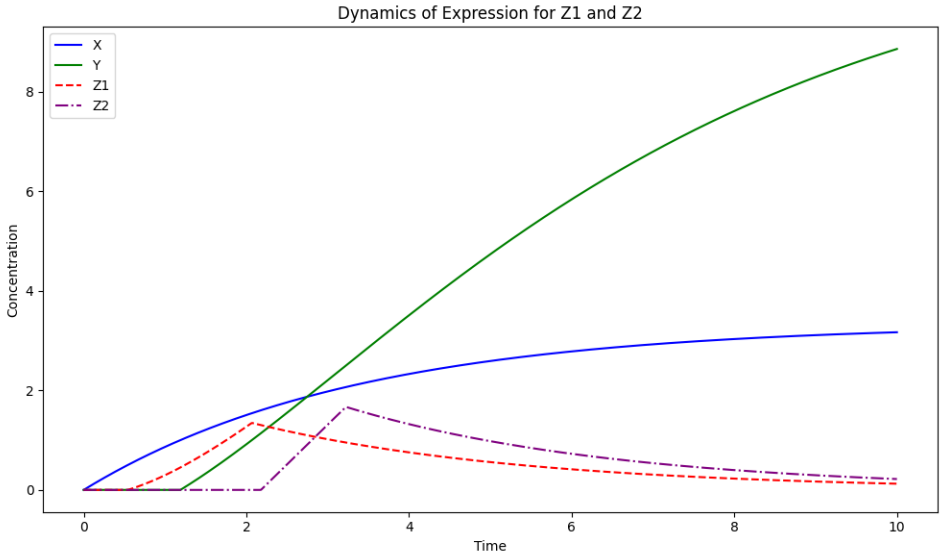

Figure 5: Simulation of Z_1 and Z_2 expression with designed thresholds

Collaboration Statement

Problem Number	**Collaborators**	**Resources**
1	N/A	Lecture 5, Textbook (PBOC2)
2	N/A	Recitation 4
3	N/A	Recitation 5, Recitation 5 Handout
4	N/A	Lectures 6-8 Notes, Recitation 5 Handout