Quantitative Molecular Biology
Problem Set #3
due: Friday, February 12, 2021

1. **Regulatory function:** The input-output characteristics of transcriptional control is described by a regulatory function $\mathcal{R}$, defined as the ratio of the transcription rate in the presence of the relevant regulatory proteins to that in the absence of the regulatory proteins. Consider the following (Hill) form of the regulatory function for a promoter controlled by a single species of transcription factor A:

$$\mathcal{R} = \frac{1 + \omega q_A^n}{1 + q_A^n}$$

where $q_A = [A]/K_A$, and n is the Hill coefficient.

(a) Plot $\mathcal{R}$ vs.[A] on both linear-linear and log-log plots for $K_A = 100$ nM, $\omega = \{3, 10, 30, 100\}$, and $n = \{1, 2, 4\}$ over the range [A] =1-1000 nM. Indicate on the plots the capacity and sensitivity of the regulatory functions.

(b) The "sensitivity" s of a promoter is defined as the maximum log-log derivative of the regulatory function. Find an expression for s in terms of the parameters ω, and n. Plot $s(\omega)/n$ vs. $\ln \omega$ for $\omega =1$-1000. Comment on the results.

2. **Inducer-controlled activation:** A transcription factor X is known to bind specifically to its target sequence with the same affinity with or without an inducer I. The effect of the inducer is to modulate the cooperative interaction between X and RNA polymerase. There is no cooperativity in the absence of the inducer. But with I bound to X, the TF-RNAP interaction is characterized by a cooperativity factor $\omega \gg 1$ when XI is bound to an appropriately located operator site.

Given a region of DNA with a weak core promoter (of strength $q_p \equiv [RNAP]/K_p \ll 1/\omega$), an appropriately located operator site for X (whose affinity is characterized by an effective dissociation constant K_X), and an inducer–TF dissociation constant K_I, apply the thermodynamic model to this promoter as follows:

(a) Draw a cartoon of the relevant states of this promoter (i.e., which site is occupied by what), and write down the associated statistical weight.

(b) Work out the probability $\mathcal{P}$ for the occupation of the core promoter by RNAP in terms of the concentration of the un-induced and induced TFs ([X] and [XI], respectively).

(c) Apply the equilibrium relation between X and I, and work out how the probability $\mathcal{P}$ depends on the total TF concentration $[X]_{total}$ and the inducer concentration $[I]$. For simplicity, assume that $[I]_{total} \gg [X]_{total}$.

(d) Make a log-log plot of the relative transcription rate predicted as a function of $[X]_{total}$ (or q_X) for $\omega = 20$ and the following concentrations of I: $[I] = \{K_I/500, K_I/50, K_I, 50K_I\}$. Indicate on the plot the capacity and sensitivity of this promoter, as well as important values of $[X]_{total}$, e.g., the onset of activation and saturation.

3. **Regulation of the *lac* promoter:** In this problem, you are asked to estimate the parameters describing the control of the *lac* promoters of *E. coli*, by analyzing the data of Oehler et al. (EMBO J 13: 3348-3355, 1994).
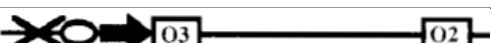
(a) The first set of data generated in the experiment of Oehler et al. was for mutants where the operators at positions O2 and O3 are both mutated completely (crosses in first column of Table below).

Fold repression, defined as the ratio of gene expression without and with the Lac repressor, was measured for two different levels of Lac tetramers ($[R] = 50$ nM and $[R] = 900$ nM) for 3 strains of cells whose operator site at position O1 bears the sequence of the operators O1, O2, O3, respectively. (The CRP site is indicated by the ellipse. For this problem you may ignore its effect.)

Promoter	$[R] = 50$ nM	$[R] = 900$ nM
O1	200	4700
O2	21	320
O3	1.3	16

Find the effective in vivo dissociation constants K_1, K_2 and K_3 for the 3 operator sites O1, O2 and O3 respectively. Compare to the value of $K_1 = 1$ pM found by von Hippel's *in* vitro experiment.

(b) For the second set of data, the operator at position O3 is removed, while the operators at positions O1 and O2 are replaced by various combinations of operators as shown below. For each of these strains, fold-repression was measured for the two concentrations of repressors:

Promoter	$[R] = 50$ nM	$[R] = 900$ nM
O1 — O2	2300	6800
O2 — O2	360	560
O3 — O2	6.8	15

Now, looping can be an issue. Using the above two sets of data to find the best estimate of the loop parameter C_{L12} (defined as in the lecture).

(c) Using the values of K_1, K_2, K_3, and C_{L12} estimated above, plot the repression factor as a function of $[R]$ on a log-log plot. Draw the data in the table in part (b) on the same plot, thereby seeing how the prediction of the thermodynamic model with DNA looping compares with the observed data quantitatively. Explain what differ in the 3 regions you see in the log-log plot. Under physiological conditions *in vivo*, the Lac repressor is expressed at ~10 tetramers per cell. Which region is the natural system in and what is the expected fold-repression in conditions with no or saturating amount of IPTG (inducer of Lac repressor)?

4. **Transcriptional AND gate:** Suppose a gene g is regulated by two TFs, TF1 and TF2, binding to operators O1 and O2 respectively. O2 is located between O1 and the core promoter. The effective dissociation constants of the operators and RNAP are K_1, K_2 and K_p. TF1 and TF2 bound to O1 and O2 interact with each other with a cooperativity factor ω_{12} and TF2 bound to O2 interacts with promoter-bound RNAP with a cooperativity factor ω_{2p}. There is no interaction between an activator bound to O1 and the RNAP.

Below, we study which parameter values would provide AND gate behavior.

(a) Write down the truth table for a Boolean AND gate. Describe how the promoter of g could act as a transcriptional AND gate.

(b) Write down the expression for the probability $\mathcal{P}$ of promoter occupation by RNAP assuming that the thermodynamic approximation holds. Assuming further that $q_p = [\text{RNAP}]/K_p \ll 1$ and $\mathcal{P} \ll 1$, derive the regulatory function $\mathcal{R}$ and write it in terms of $q_1 = [\text{TF1}]/K_1$, $q_2 = [\text{TF2}]/K_2$, ω_{12} and ω_{2p}.

(c) Along the line $[\text{TF1}] = [\text{TF2}]$, $\mathcal{R}$ increases from 1 (at $[\text{TF1}] = [\text{TF2}] = 0$) to some maximum value ω. To have a good AND gate, this transition must be steep. Show that a Hill coefficient of 2 can be achieved if $\omega_{12} \gg \omega_{2p} \gg 1$. I.e., along the diagonal $[\text{TF1}] = [\text{TF2}]$, $\mathcal{R}$ can be written in the form

$$\mathcal{R} \approx \frac{1 + \omega \left([\text{TF1}]/K\right)^2}{1 + \left([\text{TF1}]/K\right)^2}$$

with $K = K_1\sqrt{r/\omega_{12}}$, provided that $r = K_2/K_1$ is in the range $1 \lesssim r \lesssim \omega_{2p}$.

[Hint: Generally, the numerator and denominator of $\mathcal{R}$ are polynomials of the form $1 + ax + bx^2$. The Hill form $1 + bx^2$ is approximately obtained if the linear term ax is small compared to 1 and to bx^2 for the entire range of x. This algebraic exercise can be skipped by those from the more biological background.]

(d) What should K be if we want the transition to be at $[\text{TF1}] = [\text{TF2}] \approx 300$ nM?

(e) If TF2 is present at high concentrations (~ 1000 nM) and $[\text{TF1}] = 0$, the expression of the transcriptional AND gate should be low ($\mathcal{R}$ small). What is the best choice for r given the constraints demonstrated in (c)?

(f) Assume $\omega_{12} = 150$ and $\omega_{2p} = 20$; use the values you just selected for r and K to compute the corresponding K_1 and K_2.