

MCB137L/237L: Physical Biology of the Cell  
Spring 2020  
Homework 9:  
(Due 4/16/20 at 3:30pm)

Hernan G. Garcia

**1 A Feeling for the Numbers: Microbes and Viruses as the Unseen Majority**

One of the key arguments that we have made throughout the course is that sometimes just having a feel for magnitudes is a useful guide to intuition. Indeed, our model building has usually followed the sequence: simple estimates and feeling for the numbers, simple toy models, more realistic models. For this problem, we follow up on our discussions in class about just how much we don't know about the diversity of bacteria and viruses by reading the article entitled "Prokaryotes: The Unseen Majority" as well as two papers on HIV, and using them as the basis for your estimates.

(a) Throughout the course, we have used *E. coli* as our biological standard ruler. This cell has hall of fame status in biology and it is important that you have a sense of what these cells are like. Justify the assumption that a typical (i.e. *E. coli*) bacterial cell has a volume of  $1 \mu\text{m}^3$ . Also, express this volume in femtoLiters. The claim is made (see the paper on prokaryotes as the unseen majority on the course website) that in the top 200 m of the world's oceans, there are roughly  $10^{28}$  prokaryotes. Work out the total volume taken up by these cells in  $\text{m}^3$  and  $\text{km}^3$ .

(b) Recall that roughly 2-3 kg of bacteria are to be found in the waste factory of your large intestine. Make an estimate of the total number of bacteria inhabiting your intestine and then all of the intestines of all of the humans currently on the Earth.

(c) Bacteriophage are the viruses that infect bacteria. Given that the concentration of phage is tenfold or more higher than that of bacteria, report the concentration of phage in the ocean in mg/mL. Then, use this to make an estimate of the total number of phage on the Earth.

(d) Estimate the number of protein units that make up a viral capsid for influenza virus. In addition, estimate the number of lipid molecules associated with one of these viruses. The lipid molecules surround the protein coat in lipid bilayer form. Make sure you show a picture of the virus and give a rough description of what the structure is like - where is the nucleic acid, what is the shape, etc.. Also, describe the genome of this virus and compute

the total length of the nucleic acid molecules if they were strung together one after the other.

(e) Use Fig. 1 of the paper by Briggs *et al.* on HIV that is posted on the course website to measure the dimensions of the immature HIV virion. Based on what you know about the makeup of the virion, estimate the number of lipids and the number of Gag proteins in the immature virion. Then, consider the mature virion and using Fig. 2 of the second paper by Briggs *et al.* estimate the number of proteins in the ice-cream-cone-shaped capsid. How does your estimate for the number of proteins in the capsid compare to your estimate for the number of Gag proteins?

## 2 Simple repression

In class, we derived a mathematical model of how simple repression dictates gene expression levels. In particular, we showed that the fold-change in gene expression is given by

$$\text{fold-change} = \frac{1}{1 + \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon}}, \quad (1)$$

where  $R$  is the number of repressors,  $N_{NS}$  is the number of non-specific sites, and  $\Delta\varepsilon$  is the repressor binding energy. Experimentally, we can create bacterial strains that express a YFP reporter under the control of the repressor. If the YFP fluorescence signal is  $F_{reporter}$ , then the fold-change is measured by calculating

$$\text{fold-change} = \frac{F_{reporter}(R)}{F_{reporter}(R=0)}. \quad (2)$$

However, there is an extra subtlety that has to be taken into account when obtaining such fluorescence measurements. The fluorescence we measure under the microscope is the total fluorescence  $F_{total}$ . This magnitude contains the YFP reporter signal as well as the intrinsic fluorescence of the cells such that

$$F_{total} = F_{reporter} + F_{auto}, \quad (3)$$

where  $F_{auto}$  is the autofluorescence of the cell. As a result, we need to be able to subtract the cells' average autofluorescence if we want to report on  $F_{reporter}$ . Thus, the fold-change is obtained by calculating

$$\text{fold-change} = \frac{F_{total}(R) - F_{auto}}{F_{total}(R=0) - F_{auto}}. \quad (4)$$

We already learned how to extract fluorescence levels from microscopy images of bacteria using Python. In this problem, you will use the code you wrote in class in order to test the prediction made by Equation 1. You can get our final version of that code as well as the full bacterial gene expression data set from website. Note that you can also learn more about this protocol from “Computational Exploration: Extracting Level of Gene Expression from Microscopy Images” in chapter 19 of PBoC2.

(a) You can find `lacI_titration` folder under “Data Folder” shared on the course website. This data set corresponds to bacteria containing varying Lac repressor copy numbers, and a reporter containing one of three Lac repressor binding sites (called O1, O2 and O3) that control YFP expression. The repressor copy numbers for each strain are given in the following table.

Strain	Repressor number
Delta	0
R22	$22 \pm 4$
R60	$60 \pm 20$
R124	$124 \pm 30$
R260	$260 \pm 40$
R1220	$1220 \pm 160$

If you want to learn more about the measurements of the repressor copy number, you can refer to Garcia2011c, which is provided on the website. In addition, a strain called “Auto” lacking a YFP reporter is provided with the data set. This strain can be used to measure  $F_{auto}$ . Another strain called “Delta” can be used to measure  $F_{total}(R = 0)$ .

The name of the files can be interpreted as follows:

- Lac repressor binding site: O1/O2/O3
- Repressor copy number: Delta, R22, R60, R124, etc.
- Type of image: phase/yfp corresponding to the phase contrast and fluorescence snapshots of the cells, respectively.

For example, a file named “O2\_R22\_yfp\_pos\_01.tif” is a YFP image of a reporter containing the O2 Lac repressor binding site, and repressor copy number  $22 \pm 4$  (see the table above). Finally, each file name ends with “pos\_XX.tif”. This suffix corresponds to the position on the microscope slide that each snapshot was taken at. For each combination of Lac repressor binding site and number of Lac repressor molecules you will have 21 unique snapshots whose analysis you will combine to obtain the corresponding average  $F_{total}$ .

Use your code to calculate the fold-change in gene expression as function of repressor copy number by performing the various fluorescence measurements prescribed by Equation 4. Plot your results on a log-log plot.

To get a list of all the images with a given binding site and repressor number, you can write

```
#Folder where we'll get the images from
ImageFolder='/content/drive/My Drive/Colab Notebooks/data/lacI_titration/'
#Bacterial strain (combination of binding site and number of repressors) ...
  you want to analyze
BacterialStrain='O2_R22_'

# filename structure
filename_structure_phase = ImageFolder + BacterialStrain + 'phase_pos_*.tif'
filename_structure_yfp = ImageFolder + BacterialStrain + 'yfp_pos_*.tif'

# grab all images for this given strain
phase_names = np.sort(glob.glob(filename_structure_phase))
yfp_names = np.sort(glob.glob(filename_structure_yfp))
```

Here, we're getting all phase contrast and fluorescence images corresponding to the O2 binding site and 22 repressors per cell. Note that we're assuming that your working folder (where your code is saved) contains data folder that includes `lacI_titration` dataset. Please make sure to double check in your CoLab environment where your data folder is saved. You can check this by clicking the folder icon on the left corner (Files). You're going to run your code for all images in `phase_names` and `yfp_names`. For each image, you want to save the fluorescence for each cell present. As you analyze new images, you want to save your results in the vectors which we initialize below.

```
intensities=[]
```

First, you want to write a for-loop that goes through each cell in a given image, extracts its fluorescence per pixel, saves it in `cell_mean_intensity`, and appends this information to `intensities` by doing

```
intensities = intensities + [cell_mean_intensity]
```

Then, you will need to write a bigger for-loop that goes through each image in the folder and appends the extracted values of cell fluorescence intensity per pixel (saved in `intensities` as described above) into `intensities_all`. (this vector needs to be initialized before this for-loop by doing `intensities_all=[]` such that

```
intensities_all = intensities_all + intensities
```

Once you're done with processing, you can calculate the average fluorescence for the strain.

(b) Using the data you just analyzed, estimate the binding energy of O1, O2 and O3 by fitting Equation 1 “by eye” by trying a reasonable range of parameters for  $\Delta\epsilon$ . How do these compare to the binding energies obtained in Garcia2011c using an enzymatic assay instead of fluorescence as a report of gene expression? EXTRA CREDIT: Alternatively, you

can perform a least-squares minimization as discussed in class. If you want to go down this route, you might find it better to fit your data to the fold-change expressed in the language of dissociation constants

$$\text{fold-change} = \frac{1}{1 + \frac{[R]}{K_d}}. \quad (5)$$

Here,  $[R]$  is the concentration of repressors inside the cell, and  $K_d$  is the dissociation constant of repressor to the DNA.

### 3 A minimal genetic switch

In class, we introduced how genetic switches can be constructed using two repressors that repress each other's gene expression. Here, we consider a simpler regulatory architecture that can also result in a genetic switch. Specifically, we will model the self-activation of an activator molecule. For this problem, you might find it useful, to review the phase portrait concept we described in class for the case of mRNA production and degradation, as well as the phase diagram of the logistic equation you had to draw earlier on in the semester.

Figure 1(A) presents a regulatory architecture, where an activator activates its own production. Figure 1(B) shows the states and weights for our model. In this problem, we will ignore mRNA and associate the rate indicated in the figure with the rate of protein production. Here, a promoter has two activator binding sites in its vicinity. In the absence of activators, or in the presence of only one activator, the rate of protein production is  $r_0$ . When both activators are bound, the rate is  $r$ . The activators bind with a dissociation constant  $K_d$  and interact with a cooperativity factor  $\omega = e^{-\beta\varepsilon_{int}}$ , where  $\varepsilon_{int}$  is the interaction energy between activators.  $A$  is the concentration of activator.

(a) Write down an equation describing the temporal evolution of the number of activators. Consider the rate of production stemming from the model shown in Figure 1, as well as a rate of protein degradation  $\gamma$ . Hint: Remember that the overall rate of production  $\langle r \rangle$  of a system is given by

$$\langle r \rangle = \sum_i p_i r_i, \quad (6)$$

where  $p_i$  is the probability of the system being in state  $i$ , and  $r_i$  is the rate of production when the system is in that state.

(b) Plot the phase diagram for this equation in order to find how many equilibria the system can support. Namely, plot the rates of production and degradation as a function of the activator concentration. Use  $K_d = 5$  nM,  $\gamma = 0.1/\text{min}$ ,  $r_0 = 0.01$  nM/min, and  $r = 0.5$  nM/min. Make plots for  $\omega = 1$  and  $\omega = 10$ .

(c) Draw vectors indicating the direction of the concentration change under your plots as we did in class for the mRNA production and degradation case, and as you explored in the homework in the context of the logistic equation. How many equilibrium points do you find? Indicate whether these points correspond to stable or unstable equilibria. You can review the concept of phase portraits by reading “Computational Exploration: Growth Curves and

the Logistic Equation” on page 103 of PBoC2, paying special attention to vectors drawn on the lower part of Figure 3.10.

(d) Solve the equation you derived in (a) for different initial conditions, and plot all of them on the same graph. Choose initial conditions that help illustrate how the system can converge to different levels of activator in steady state.

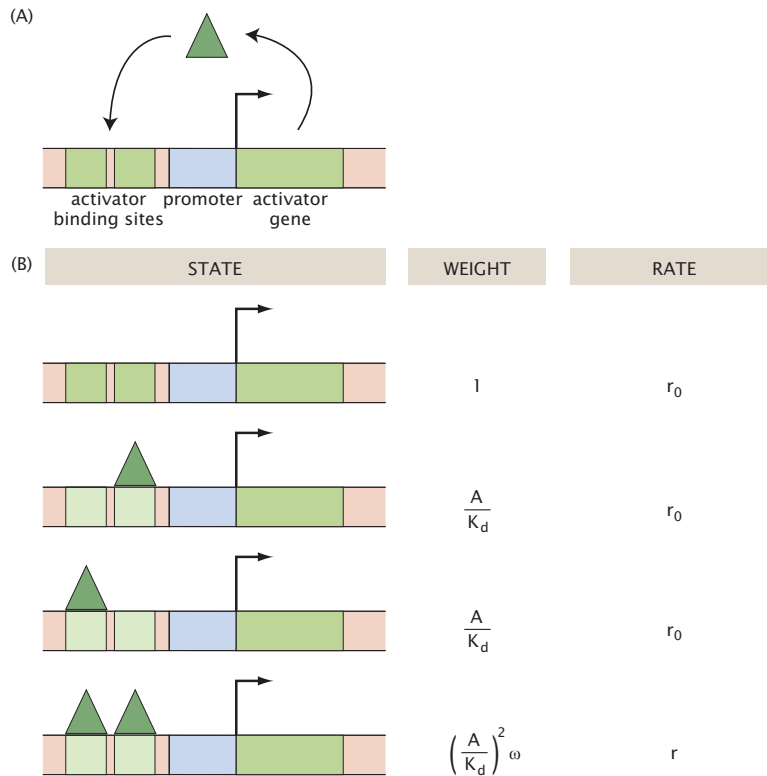


Figure 1: A simple autoactivation switch model. (A) Cartoon of the autoactivation switch. (B) States, weights and rates for the autoactivating genetic switch model.